

Efficacy Trial of FFT Tutoring with the Lightning Squad
Evaluation Protocol
Evaluator: National Foundation for Educational Research (NFER)
Principal investigator: Stephen Welbourne



Template last updated: December 2022

Evaluation summary

Project title	Efficacy trial of FFT Tutoring with the Lightning Squad, a two-armed cluster randomised trial
Developer <i>(Institution)</i>	FFT Education
Evaluator <i>(Institution)</i>	National Foundation for Educational Research (NFER)
Principal investigator(s)	Stephen Welbourne
Protocol author(s)	Lillian Flemons, Aarti Sahasranaman, Gemma Schwendel, Stephen Welbourne
Trial design	Two-arm interleaved randomised controlled trial with random allocation at school level
Trial type	Efficacy
Pupil age range and Key stage	7 – 9 years, KS2 (Year 3 and Year 4)
Number of schools <i>(at design stage)</i>	72
Number of pupils <i>(at design stage)</i>	1,152
Primary outcome measure and source	Reading skills (Woodcock Reading Mastery Test 3 rd Edition (WRMT-III))
Secondary outcome measure and source	Attitudes towards reading (Section R7 of the student survey from Progress in International Reading Literacy Study (PIRLS))

Contents

Evaluation summary	1
Protocol version history	3
Study rationale and background	4
Intervention.....	6
Theory of Change	15
Impact evaluation design.....	18
School eligibility	20
Pupil eligibility	21
Implementation and process evaluation (IPE) design	28
Cost evaluation design.....	44
Ethics and registration	45
Data protection.....	47
Legal Basis	47
Data Collection and Sharing.....	47
Personnel.....	50
Risks	51
Timeline	53
References.....	55
Appendix A: questions from section R7 of PIRLS 2021	59

Tables

Table 1: Trial design	19
Table 2: Sample size calculations	23
Table 3 Sequencing of IPE data collection activities.....	33
Table 4: IPE methods overview.....	34

Figures

Figure 1: Theory of Change for the Tutoring with the Lighting Squad Programme..... 16

Protocol version history

Version	Date	Reason for revision
1.0 [original]	22/09/2025	N/A

Study rationale and background

Reading skills serve as the foundation for pupils to access learning across all subjects in the curriculum (Department for Education, 2023). They have an impact on academic success and the ability to access wider life and employment opportunities (Marguerita and Conlon, 2016; Mulcahy, Bernardes and Baars, 2019). Research has shown the importance of getting children off to a good start early as children who are poor readers at the primary level are less likely to complete compulsory education (Hernandez, 2012). Children from disadvantaged backgrounds, those with Special Education Needs and Disabilities (SEND) and those with English as an Additional Language (EAL) also have lower reading attainment compared to their peers (GOV.UK, 2025).

Improving reading attainment for pupils is, therefore, an important policy priority for the UK government. One of the “Levelling Up Missions” of the government is to ‘significantly increase’ the number of primary school children achieving the expected standard in reading (as measured by pupils’ performance in the statutory Key Stage 2 (KS2) English reading assessment) by 2030 (HM Government, 2022). In England, the aim is for 90% of children to achieve the expected standard and for the percentage of children meeting the expected standard in the worst performing areas to increase by a third (ibid).

Achieving this mission within the next five years would, however, require a substantial amount of work. KS2 reading attainment has not improved significantly since before the COVID-19 pandemic. For instance, in 2023/24, 74% of pupils in KS2 achieved the expected standard in reading compared to 75% in 2018 (GOV.UK, 2025, p. 2), (DfE and STA, 2018). Pupils from disadvantaged backgrounds fared worse, with only 57% achieving the expected standard in reading in 2023/24 (ibid).

Tutoring could be an effective strategy to increase the reading attainment of all pupils and close the disadvantage gap. While one-to-one tutoring has been shown to have the greatest impact on attainment (Harrison and Higgins, 2023), it can be expensive to implement. Small group tutoring, where a teacher or teaching assistant (TA) works with a group of two to five pupils can bring down costs while also providing an effective strategy to improve attainment (ibid). This approach is most likely to be effective when it is targeted to the specific needs of pupils, training is provided to the tutor and the tutoring sessions are no longer than an hour and take place frequently (e.g., three times a week or so). Studies of small group tutoring focused on reading have shown an impact of an additional four months’ progress on average in primary schools (EEF, 2021b). Specifically, low attaining pupils benefit from small group tutoring (ibid). The average cost of small group tutoring is lower than one-to-one tutoring; the cost of small group tutoring programmes involving more pupils per group or online programme delivery typically tend to be lower. Staff time spent on training and delivery are also factors that influence the overall cost of the small group tutoring programmes to schools. This suggests that small group tutoring is a worthwhile approach for schools to consider to increase reading attainment, particularly for low attaining pupils.

Technology offers an additional avenue to potentially enhance the effectiveness of tutoring. Whilst computer-assisted instruction without the inputs of a human tutor have been not been found to be effective with struggling readers (Slavin *et al.*, 2011; Cheung and Slavin, 2013), there is evidence to suggest that combining human tutors with technology can have strong positive effects. In fact, technology-assisted small group tutoring approaches have been shown to improve the reading attainment of struggling readers at the primary school level (Madden and Slavin, 2017; Chambers *et al.*, 2011). This impact is likely because human tutors can support children with phonics, helping them read and sound out words and support with reading comprehension, which cannot be accomplished by the use of existing standalone technologies. The continuous monitoring provided by the technology platforms has the potential to allow human tutors to take a data-driven approach to further augment their support to pupils.

Collaborative or cooperative learning approaches involve pupils working together in pairs or small groups of three to five on learning tasks and activities. Cooperative learning approaches have been demonstrated to have consistently positive effects on pupil attainment and motivation (EEF, 2021a; Lewin *et al.*, 2024). For example, primary pupils using cooperative learning approaches make an additional five months' progress. For literacy interventions, these approaches have been shown to have an impact of an additional three months' progress on average. Although there is limited evidence on the differential impact of cooperative learning approaches for pupils from disadvantaged backgrounds, there is some evidence to suggest that these approaches can benefit low attaining pupils by allowing them to work with their peers to share knowledge, skills and articulate thinking (*ibid*). While cooperative learning can take on many different approaches, the most effective cooperative learning takes place when structured approaches with well-designed tasks are used and talk and interaction between learners is promoted (*ibid*).

Tutoring with the Lightning Squad (TWL), developed and delivered by the Fischer Family Trust (FFT), uses a blended learning approach where face-to-face tutoring by a human tutor is supported by an online tutoring platform. Pupils identified to be behind age-related reading expectations work in pairs on the online platform with the support of a trained tutor to improve their reading skills. Each tutor supports a group of four pupils (i.e., two pairs) in each tutoring session. Pupils work through stories and the learning tasks; within each story level are structured and designed activities to improve word reading (phonics), encoding, fluency and comprehension.

An internal analysis by FFT of data collected from pupils who used the TWL programme in the 2020-21 academic year shows the programme's promise in improving reading outcomes (Bibby *et al.*, 2023). Among pupils who received 20 weeks of tutoring, the median WCPM (words correct per minute, a measure of oral reading fluency) improvement was 9 words in 20 sessions. This is more than three times the rate of improvement implied by the benchmark of 0.72 words per week during the first four years of schooling. The progress rate in WCPM is similar for pupils receiving a larger dosage of tutoring (i.e., 30 weeks). The report noted issues with the quality of data collected for individual pupils (e.g., noise in data on progress in decoding) although this was mitigated to some extent by the large dataset used for the analysis. However, since the study only

assessed improvements among pupils using TWL, the lack of a control group precluded any conclusions about efficacy of the programme.

The impact of TWL on reading outcomes for pupils has previously been evaluated using a randomised controlled trial (RCT) (Ross, Laurenzano and Madden, 2017). This trial, conducted among first, second and third grade students (i.e., 6 – 9 years old) in six elementary schools in the USA, showed that TWL produced a significant positive impact on the Word Attack (measure of decoding skills) and Passage Comprehension subscales of the Woodcock Reading Mastery Test (ibid). The trial included 165 children, and the analysis only included intervention pupils who had complied with the intervention (defined in the trial as having completed at least 25 sessions of TWL).

This efficacy trial will be the first to evaluate the impact of the TWL programme in English schools using a robust experimental design. This RCT is larger in scope than the previous study and will involve 1,152 pupils in Years 3 and 4 (ages 7 - 9) in 72 primary schools in England who will receive access to the TWL programme for around 16 weeks. Schools will be randomly assigned to either receive the intervention in Year 3 (arm 1) **or** in Year 4 (arm 2). The year group not receiving the intervention in each arm will serve as the control group for the year group receiving the intervention in the other arm. Compared to a traditional school-randomised design, this interleaved design gives better statistical power. We anticipate that when the ICC is split between year-group and school levels most of the variance will be accounted for at year-group level. This means that each extra year-group within a school is nearly as good as adding another school in terms of statistical power. A second advantage of this design is that it will be much more attractive to schools with all schools receiving the intervention. We anticipate that this will greatly facilitate recruitment. Pupils will receive access to the TWL programme during the 2025-26 academic year. The primary outcome is reading attainment measured using the Woodcock Reading Mastery Tests 3rd edition (WRMT-III). The secondary outcomes include decoding, oral reading fluency and reading comprehension as assessed by individual subscales of the WRMT-III. An additional secondary outcome will be pupils' attitudes towards reading as measured by the Section R7 of student survey from the Progress in International Reading Literacy Study (PIRLS). The intention-to-treat (ITT) approach will be used for all primary and secondary analyses and the Complier Average Causal Effect (CACE) analysis will be used to assess the impact of the programme for pupils who comply with the intervention. The integrated Implementation and Process Evaluation (IPE) will seek to explore how and in what circumstances TWL impacts pupils and teachers. We will explore these questions using a combination of qualitative and quantitative methods. The qualitative methods will consist of interviews with teachers and senior leaders and focus groups with pupils while the quantitative methods will involve analysis of surveys of teachers and analysis of TWL platform data. The cost evaluation will provide a robust estimate of the cost to schools per pupil per year of delivering the TWL programme.

Intervention

The Tutoring with the Lightning Squad (TWL) intervention will be implemented for the purpose of this efficacy trial between November 2025 and May 2026.

A detailed description of the intervention in the context of the TIDieR framework is presented below.

Why: Rationale, theory and/or goal of essential elements of the programme

TWL initially aims to improve pupils' decoding in tandem with fluency. Once accuracy is mastered, the focus shifts to automaticity with longer texts. This approach is congruent with the latest theoretical ideas on learning to read, which emphasise phonological decoding skills as being essential to the early stages of learning to read with this being supplemented with a direct mapping between words and meaning as children become more expert (Castles, Rastle and Nation, 2018). Reading is the gateway through which all education is accessed, which is why reading proficiency is associated with attainment across the curriculum (GL Assessment, 2020).

The TWL programme was designed in response to existing evidence on effective interventions. The Education Endowment Foundation (EEF) cites tutoring as one of the most effective ways of bridging attainment gaps. Moreover, they recommend one-to-one or small group tutoring, with frequent sessions of up to an hour, multiple times a week for at least six weeks. The EEF has also highlighted the importance of responsive tutoring, where specific learning gaps are identified through diagnostic assessment and addressed through intensive targeted support.

Building TA subject knowledge through strong training, supported by detailed resources has also been found to enhance the impact of interventions (Sharples, Webster and Blatchford, 2021).

Finally, the EEF has highlighted the benefits of both collaborative learning and the use of digital technologies, supported by meta-analysis evidence indicating positive effects for computer-supported collaborative learning (Chen *et al.*, 2018).

The programme content and activities link to the National Curriculum objectives for reading, spelling and comprehension. KS1 common exception words and words from KS2 Spelling Lists are also covered.

The core components¹ of the TWL programme constitute:

- Small groups of pupils working with the Tutor at any one time
- Frequent tutoring sessions (that is, multiple times per week)
- Cooperative learning within the pupil pairs
- Hybrid support from both the Tutor and TWL platform
- Learning opportunities tailored to the child's level of reading attainment

¹ That is, 'those features that are expected to influence the intervention's success if implemented successfully and without which the intervention would be expected to be less effective' (EEF, 2022a).

Who: Recipients of the programme

The programme is designed for pupils from Reception to Year 6 with a reading attainment below what would be expected for their age. However, since this trial is focused on establishing efficacy of the programme, the participating year groups will be restricted to Year 3 and Year 4.

For the purpose of this trial, teachers will use their professional judgement to nominate half of their cohort from each of the focus year groups for testing at the end of the start of academic year 2025/26. This group should constitute the pupils with the lowest reading attainment in their cohort. The nominated pupils will complete FFT's Reading Assessment Programme (RAP).² Teachers will then use the RAP scores and their own professional judgement to identify eight pupils per year group to participate in the trial (as either treatment or control). FFT will conduct a training session ahead of the pupil nomination process to provide teachers with guidance on administering the RAP and nominating and selecting pupils. Briefly, FFT guidance stipulates that pupils must have at least a basic decoding ability (e.g., able to read the word 'cat') to participate in the TWL programme. Similarly, pupils with severe special educational needs and/ or disabilities (SEND), severe visual impairment and selective mutism, may not be able to access the programme. Teachers should also keep in mind that it will be necessary to constitute four pairs of pupils in each year group with similar levels of reading attainment. A RAP threshold will be stipulated to ensure that pupils with very low reading attainment are not included in the trial.

Pupils receiving the intervention will complete an initial assessment on the TWL online platform to establish the story level at which they should begin the programme, as well as to identify areas of strengths and weaknesses. Pupils participate in the programme in teacher/TA-allocated pairs, based on similar reading attainment and perceived compatibility. Pupils within a pair should be no more than five story levels apart.

In usual practice, pupils can be paired across year groups depending on reading level. However, this will not be possible within the context of the evaluation as there will only be one year group per school participating in the programme.

Pairs can be dissolved and/or re-paired during programme delivery, or pupils can work individually, if incompatibility is observed – whether due to speed of progress, behaviour or learning styles. In the context of the trial, re-pairing can only occur among pupils receiving the programme i.e. pupils in the control group cannot be paired with pupils receiving the programme.

What: Materials

Training

Training is delivered via a one-day (six-hour) live online training session. It is recommended that the School Coordinator (SCs) and two other staff members (teachers or TAs) attend the training.

² [Reading Assessment Programme - FFT](#)

However, a minimum of two staff members is accepted. Between 15 and 30 participants would be expected for each training session.

As a number of training slots will be made available for the trial, there is flexibility for school staff to attend another session if they miss the one that they had initially registered for. FFT will run catch-up sessions in any case where a school is unable to attend any of the sessions initially made available.

In addition to the online training, schools have optional access to 16 short videos covering specific aspects of the programme that they can refer back to at any time.

Online training sessions about administering the RAP are made available and completion of this training is required for the trial. SCs (and any colleague(s) who will be administering the RAP assessment) will be required to attend the 45-minute briefing about how to administer the baseline assessment. In usual practice, completion of the RAP training is recommended but not mandatory.

Session delivery

Access to the TWL platform is required for all sessions. Schools require a device per pupil pair, with high-quality speakers either embedded in or in addition to the device. FFT can provide schools with laptops if they do not have the necessary technology available. Internet connection must be available within the allocated space – FFT will support schools to check the strength of the connection before session delivery starts.

A number of physical resources are also provided for use during the sessions. Schools receive 60 copies of FFT's Further Reader texts. Pupils are provided with a whiteboard for writing words on and cards outlining the roles of Reader and Coach. The Tutor may draw on additional support materials, including physical copies of the texts and activities, and are provided with certificates and stickers to celebrate pupil success. Parent Progress reports that outline the words a pupil has recently mastered can be shared with parents.

Tutors are provided with a Support Guide that provides resources and information to help them support pupils during the sessions, ensure pairs are working effectively together, link the sessions to the school's approach to reading and phonics, and monitor pupils' progress to identify any additional support needed.

The Scope and Sequence resource, supplied by FTT, provides schools with the structure, content and learning focus for each of the stories. In addition, the User Guide provides detailed guidance on using the TWL platform.

There are 10 interactive automated analytic reports available on the TWL platform to track pupil progress at school, year group and pupil level, as well as use of Independent Access, session

attendance and Tutor Check outcomes.³ Further information about these elements is provided in the following section. Schools are provided with guidance for interpreting the reports.

What: Procedures, activities and/or processes

Training

The online training session for SCs and teachers/TAs covers:

- Logging on and getting started on the TWL platform
- The six activities and Tutor Check (see below for more detail)
- Top tips for tutors
- Resources
- The role of the Tutor
- Analytic reports
- Phonics knowledge

The last two components are usually optional but are mandatory for the purpose of the trial.

Session delivery

Each programme session involves one Tutor supporting two pairs of pupils to work on the TWL platform.

To start, participating pupils complete the TWL assessment with the Tutor. This assesses their decoding and fluency level and will provide a recommendation for the story level at which the pupil should start the programme. This assessment takes approximately five minutes per pupil. The assessment is adaptive to pupil performance, and pupil performance in the decoding element determines the difficulty level of the passage provided for the fluency task. The platform then recommends at which level the pupil should start, although it is possible for the Tutor to override this if they are aware of gaps that the assessment may not have identified. Each pupil is paired with another pupil of a similar reading attainment and perceived compatibility for collaborative learning. The system automatically assigns the pair the lower of the two assessed story levels. The two pupils then progress through the levels as a pair. For this trial, schools will choose a maximum of eight pupils (that is, four pupil pairs).

There are 120 levels available on the TWL platform, each with a unique story written especially for the programme. Stories can be fiction, non-fiction or poetry and increase in difficulty as the pupils advance through the levels.

Each level contains four activities based on the story and associated vocabulary:

³ The 10 available reports are: School Summary, Pupil Summary, School Progress, Pupil Progress, Activity, Activity Item: School, Activity Item: Pupil, Pupil Attendance, Tutor Check Pupil, Independent Access, and Parent Progress.

- Power Reader: pupils practise reading stories together to develop fluency
- Super Speller: pupils practise spelling words
- Word Quest: pupils practise the vocabulary associated with the story
- Story Quest: pupils read the story out loud and answer comprehension questions to reinforce their learning

Two additional activities are also available for readers working at the expected level for Year 1 or Year 2:

- Word Blender: pupils practise decoding individual words
- Letter Launch: pupils vocalise a set of phonemes to improve knowledge of grapheme-phoneme correspondence (GPC)

The pupils can choose to complete the activities in any order, with the exception of Story Quest, which only becomes available once the other five activities have been completed to the required level (see below). Pupils are not expected to complete all six activities in each session.

Pupils complete the activities with a partner taking turns as the Reader and Coach, swapping roles when indicated by the platform. The Tutor explains each of the roles at the start of the sessions. There are also child-friendly 'role cards' that pupils can refer to as a guide. The Tutor models and provides feedback and specific praise around cooperative learning behaviours.

The programme provides clues to help the pupil work out the answers for themselves, and the pupils mark each other's answers (or their own if working individually). Children must achieve 50% in Word Quest, 80% in Power Reader and 75% in all other activities for them to be marked as complete. Children who are 'stuck' on a particular activity or story level should receive additional teaching support from the Tutor until they are ready to move on to the next level. Guidance for providing this support is covered in the training and in the Support Guide for Tutors, although these constitute suggestions for further scaffolding approaches rather than a specific schedule of exercises. It is ultimately up to the Tutor to decide how best to support each pupil.

Once the pair have completed the Story Quest, the Tutor will complete the Tutor Check with each pupil in turn to ensure they are ready to move on to the next level. Both pupils in the pair must pass the Tutor Check in order to progress to the next story level. The Tutor Check involves decoding words from the story, reading a passage from the story for fluency and answering a comprehension question about the passage. These checks are used by the system to track pupil progress, alongside the regular re-assessment of pupils (see below). Tutors are provided with guidance on determining when a pupil is ready to pass, informed by the Tutor Check score.

Pupils receive reading-related rewards when they pass a Tutor Check and earn bonus points that they can trade for a video or game related to reading. Certificates can also be printed off the platform and presented to the children. If a pupil does not pass the Tutor Check, the Tutor uses their professional judgement to decide whether both pupils in the pair would benefit from repeating the story level or whether to move both pupils onto the next story and deliver some additional direct teaching on the word(s)/elements the pupil was struggling with if it is significant.

Another option would be to re-pair the pupils if the Tutor felt that they were working at different levels.

The Tutor may choose to change the story level if the pair is finding the activities either too difficult or too easy. If one pupil in the pair is progressing faster than the other, it is recommended that the pair be dissolved and either re-paired or that the pupils each work individually.

The Tutor also has a direct teaching role to enhance the impact of the sessions, particularly when children are making mistakes, getting stuck, working very slowly or show other indications of finding something challenging. Strategies Tutors can use in these cases are provided in the Tutor Support Guide, including focusing on the area of difficulty through either the platform or the pupil worksheets.

Every 20 sessions (approximately every 6-8 weeks), Tutors will be prompted to re-assess the pupils to ensure they are still working at the appropriate level for both pupils. The Tutor provides encouragement, support, feedback and direct teaching to support the pupils' progress.

If one child in the pair is absent, the other does Free Play. This involves choosing a previous story and practising the activities for it. This should only be done for one or two sessions at most. In the case of longer or regular absences, the pair should be dissolved with one or both pupils entering a new pairing or working individually.

FFT usually recommends that each session involves a maximum of two pairs for KS1 pupils, and a maximum of three for KS2 pupils. For this efficacy trial, FFT recommends two pairs per session. More than this number will result in less amount of teaching time being accorded to each pair.

Independent Access offers pupils the opportunity to read three earlier stories and six reading activities to consolidate and embed their learning. Independent Access Guidance is provided for parents to help them support their child's learning. As Independent Access will not be covered in the training for the trial, it is possible but unlikely that schools will use it.

Who: Programme providers/implementers

The Tutor is responsible for delivering the sessions. They are usually experienced teaching assistants (TAs) or qualified teachers with experience of working with primary-age pupils and with a strong understanding of how primary schools teach reading and phonics. Usually, two or three Tutors are trained per year group to cover delivery in the event of absence or staff turnover. The platform allows for multiple Tutors to work with the same pairs. FFT recommends that the Tutor be familiar with the pupils already (e.g. from supporting them in class) and/or have experience developing reading skills, but Tutor selection is at the discretion of the school.

For the purpose of the trial, up to two staff members per school will complete the Tutor training. Another staff member can complete the Tutor training at a later point if needed in cases of staff turnover.

The nominated SC supports the Tutor with implementing the programme, including ensuring Tutors are given the necessary space to work with the pupils, and that pupils are permitted to leave the classroom. It is recommended that the SC be a middle or senior leader at the school.

Both the Tutor and SC must have completed their allocated training prior to implementing the programme. Ideally the Tutor and SC roles should be filled by different people.

FFT provides ongoing technical and pedagogical implementation support to schools, including weekly reviews of school data to alert SCs of any potential issues around pupil progress.

How: Mode of delivery

Training for the SC and Tutors is delivered live, online, using Zoom.

Session delivery is hybrid – the digital platform guides the session, but pupils receive in-person support from the Tutor alongside it.

Where: Location of the programme

FFT recommends that the sessions take place in a quiet room with internet access and sufficient space for two pairs of pupils to be seated so as not to distract each other but close enough for the Tutor to be able to oversee both pairs at once. It must be possible to connect to the school internet from within the space.

Schools may choose to allow pupil access to the platform at home through an Independent Access weblink.

When and how much: Duration & dosage

Subscription to the platform allows for Tutors to attend as many training sessions as they wish if at any point they want a refresher session.⁴ Training for additional staff will be provided in cases of staff turnover.

FFT recommends that each programme session lasts approximately 30 minutes. Pupils should receive as many sessions as required for them to attain the expected reading level for their age. The platform will notify the Tutor when this level is reached. Sessions should occur daily, or as regularly as possible.

For the purpose of the trial, each pupil should receive a total of between 50 and 60 sessions.⁵ FFT recommends that staff deliver a minimum of four sessions per week for at least 16 weeks. Note that the trial timetable allows for up to 24 weeks of programme delivery, providing some buffer to account to for tutor illness and other important school activities.

⁴ The platform subscription has been substantially subsidised for the purpose of the trial.

⁵ Attendance is logged when the pupil's name and avatar are selected, and they start an activity. This constitutes a session being recorded against the pupil.

FFT provides example timetables for delivery and advises schools to ensure pupils are not always missing the same lessons. It is also recommended that the pupils do not miss core lessons if they are able to access these.

It is not expected that Tutors will spend any time on the programme outside of the session time, with the exception of monitoring and sharing the data, if this is the arrangement with their school (the SC and/or the Senior Leadership Team (SLT) may do this themselves).

Tailoring: Adaptation of the programme

The platform has accessibility settings to support learning for visually impaired pupils and pupils with dyslexia. Tutors may also work with one pair or individuals where necessary, and pupils can work as an individual or pair depending on their needs. Where a pupil is struggling, the Tutor is encouraged to draw on the paper resources to provide additional tailored support to the pupil using their own professional judgement.

In exceptional cases, pupils may work as individuals if they are not able to work effectively in pairs and/or there is no suitable pupil to pair them with. The Tutor would then take on the role of Coach for that child (see 'What: Procedures, activities and/or processes').

Unacceptable adaptations to programme implementation include:

- delivering only one or two sessions per week, as the programme is designed to be delivered in short frequent sessions; and
- working with more than four pupils (two pairs) per session, as this would make it challenging for the Tutor to provide the necessary level of support.

How well (planned): Strategies to maximise effective implementation

The programme is designed to be fun, multi-sensory and pupil-driven to encourage pupil engagement. The platform is designed to be user-friendly and self-explanatory.

A seven-person customer support team is available to support schools with advice by phone or email, covering both technical matters and the teaching of literacy. Each school has a dedicated account handler as their direct point of contact for any issues or questions.

How well (actual, based on previous trials): Evidence of implementation variability

There is no previous evidence on where and when sessions delivered, how children were paired or managed or the concentration of tutoring (weekly dosage and total number of sessions). However, there are currently over 400 schools using the platform with a staff member as the Tutor and they are consistently delivering the intervention to small groups of pupils three to five times per week.

Theory of Change

The ToC for the TWL programme is shown in Figure 1. It outlines the target population, activities, outputs and short-term outcomes that are expected to lead to improved overall attainment for participating pupils. Separate strands of outputs and outcomes outline anticipated pathways for impact on parents and teachers/TAs (through the Tutor role). The following description of the ToC focuses on the aspects of greatest relevance to the IPE for this trial.

An important part of the programme is ensuring that pupils are getting tailored support at the right level for their current attainment. This means that a key part of the programme activities comprises identifying those pupils who will benefit most from the programme, ensuring that they are starting from the right story level and being paired with a partner with an attainment level that is sufficiently similar for mutually beneficial collaborative learning to occur. The ToC also highlights the fact that this is a hybrid programme, where the Tutor's professional input plays a key role in enabling pupils to engage effectively with what the online programme sessions offer.

The primary route for achieving impact on pupil reading skills is through the sessions providing pupils with the opportunity to practise a range of the necessary skills, including fluency and decoding, in the form of the platform content complemented by Tutor scaffolding and support. The role of the collaborative learning element presents a complementary route, with the support each pupil provides to their partner offering the opportunity for a more in-depth learning (EEF, 2021a)

Evidence that decoding and fluency skills predict reading comprehension (Decker *et al.*, 2018) is reflected in the mediating role they are allocated in the ToC, although all three areas constitute key outcomes of interest for the impact evaluation. The role improved reading skills can have in improving overall attainment is also noted in the long-term outcomes (Savolainen *et al.*, 2008).

In addition, the ToC highlights the role pupil enjoyment of, and confidence in, reading can play in improving reading outcomes (Education Standards Research Team, 2012). The platform is designed with the intention of encouraging positive attitudes to reading, including use of a rewards system and clear, child-friendly progress markers.

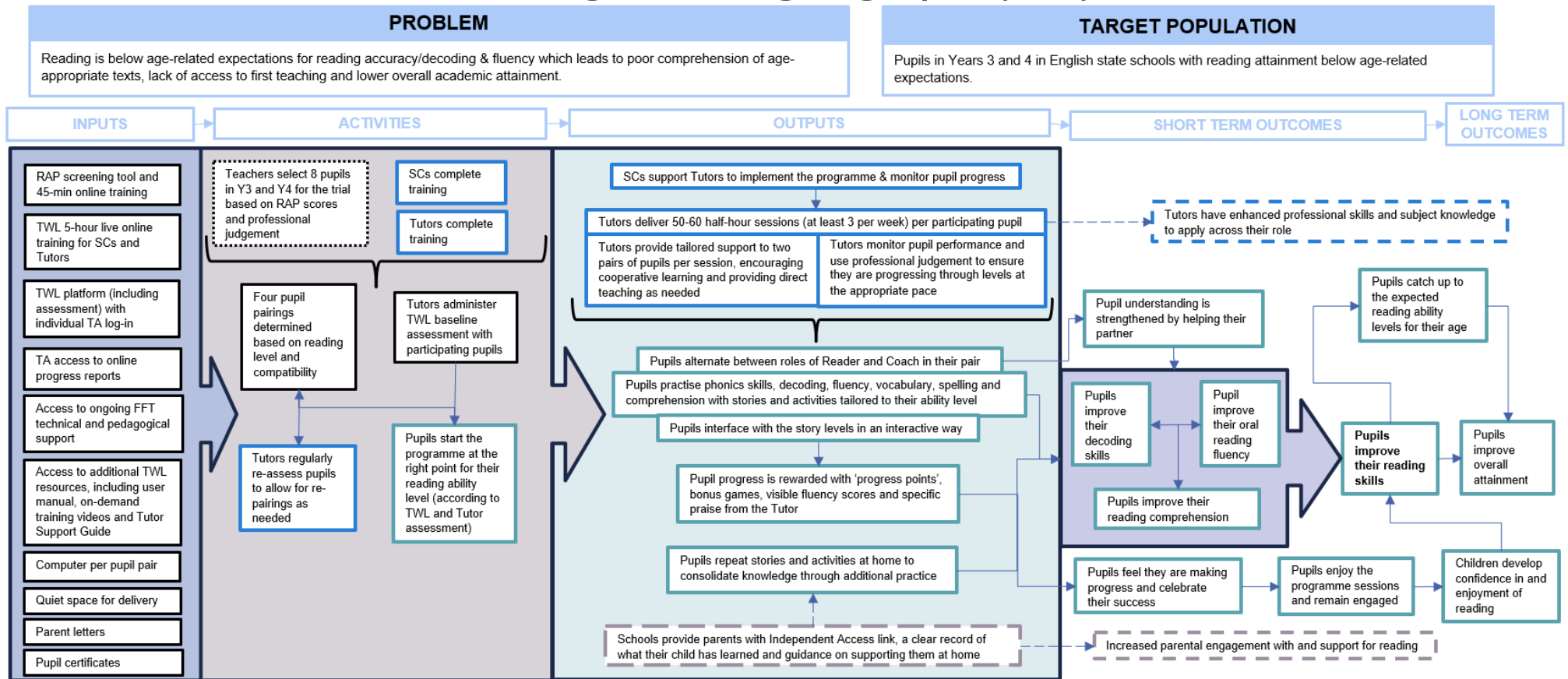
The Independent Access element of the programme has not been tested either for usage or impact. As it will not be covered in the training for the trial, use of it is unlikely to be widespread, but the IPE will seek to understand extent of usage and perceived impact where this does occur.

Teacher/TA outcomes are likewise not a focus of the impact analysis and the training they receive is primarily centred on delivering the programme effectively rather than broader professional knowledge or skills development. For this reason, it is not expected that teacher/TA outcomes will act as an important mediator for any pupil outcomes observed. However, it is likely they will still draw some professional benefit in terms of learning to better support pupils' reading development that they will be able to apply more broadly across their role. The IPE will aim to understand this further.

Finally, the ToC outlines some of the hypothesised moderators of any impact observed, including pupil characteristics such as eligibility for Free School Meals (FSM) (Ashraf *et al.*, 2021), school-level factors like staff turnover (Gibbons, Scrutino and Telhaj, 2018) and the nature of programme implementation, primarily relating to dosage and fidelity. This list is not intended to be exhaustive but represents a starting point for the IPE in terms of considering contextual factors that may require exploration and consideration.

Figure 1: Theory of Change for the Tutoring with the Lighting Squad Programme

Tutoring with the Lighting Squad (TWL)



Moderators / contextual factors

Pupil-level factors: FSM status; SEN; EAL; home access to and support with IT; IT skills; reading confidence; reading enjoyment; home reading environment and support

School-level factors: SC/TA's existing knowledge, experience, skill and commitment to TWL; SLT and broader staff support; human, physical & IT resources; level of staffing consistency

Implementation-related factors: number and frequency of sessions delivered; length of sessions; level and nature of TA support for pupils (including effective use of resources); level and nature of SC support for TAs; level and nature of FFT support; training completion; implementation fidelity; nature of learning displaced by the sessions; working alone or in a pair; quality of pairing; ability of school to foster home-school links for reading; suitability of pupils nominated for the programme

***Note:** Boxes with a dotted border represent elements that have been introduced for the purpose of the trial and would not occur in usual practice. Boxes with a dashed border do not fall within the scope of the impact side of the trial but will still be investigated via the IPE. Black borders indicate set-up, blue borders indicate teacher activities/outputs/outcomes, blue borders indicate pupil outputs/outcomes and grey borders indicate parental outputs/outcomes.*

Impact evaluation design

Research questions

The primary research question is:

RQ 1. What is the overall impact of TWL on the reading skills of children who are below expected reading levels in years 3 and 4?

The secondary research questions are listed below:

RQ 2. What is the impact of TWL on the reading skills of the following subgroups of children who are below expected reading levels in years 3 and 4?

- a. Disadvantaged children (FSM)
- b. Pupils with English as an Additional Language (EAL)
- c. Pupils with Special Education Needs and Disabilities (SEND)?

RQ 3. What is the impact of TWL on the reading skills of children who are below expected reading levels in each individual year group?

RQ 4. What is the impact of TWL on the following reading outcomes of children who are below expected reading levels in years 3 and 4?

- a. Oral reading fluency outcomes as measured by the oral reading fluency subtest of the WRMT-III.
- b. Decoding skills as measured by the Word Identification and Word Attack subtests of the WRMT-III.
- c. Reading comprehension as measured by the passage comprehension subtest of the WRMT-III.

RQ 5. What is the impact of time spent using TWL in school (measured as number of sessions) on reading skills of children who are below expected reading levels in years 3 and 4?

RQ 6. What is the impact of TWL on the attitudes towards reading of children who are below expected reading levels in year 4?

Design

Table 1: Trial design

Trial design, including number of arms		Two-arm, interleaved, four-level (pupil, class, year group, school) cluster randomised efficacy trial
Unit of randomisation		School
Stratification variables (if applicable)		N/A
Primary outcome	Variable	Reading Skills
	Measure (instrument, scale, source)	Woodcock Reading Mastery Tests 3 rd edition ((WRMT-III (Woodcock, 2011)): a derived composite score from four WRMT-III subtests using a Principal Components Analysis (PCA)
Secondary outcome(s)	Variable(s)	1. Oral reading fluency, reading comprehension and decoding. 2. Children's attitudes towards reading (Year 4 pupils only)
	Measure(s) (instrument, scale, source)	1. Four subtests from WRMT-III: a) oral reading fluency sub-test, b) passage comprehension sub-test c) Word Identification sub-test d) Word Attack sub-test 2. Eight items from section R7 of the student survey from Progress in International Reading Literacy Study (PIRLS) (IEA, 2020): total score across all eight items
Baseline for primary outcome	Variable	Reading Skills
	Measure (instrument, scale, source)	Raw score from FFT's Reading Assessment Programme (RAP) ⁶ that measures words correct per minute.
Baseline for secondary outcome	Variable	1. Baseline reading skills 2. Baseline attitudes towards reading
	Measure (instrument, scale, source)	1. Raw score from the Reading Assessment Programme (RAP) that measures words correct per minute. 2. Eight items from section R7 of the student survey from Progress in International Reading Literacy Study (PIRLS) (IEA, 2020): total score across all eight items

⁶ <https://fft.org.uk/literacy/reading-assessment-programme/>

This is an efficacy trial designed to evaluate the impact of the FFT Tutoring with the Lightning Squad programme (TWL) over one academic year. Randomisation will be at the school-level; schools will be randomised in a 1:1 ratio into two arms: schools in arm 1 will deliver TWL to pupils in Year 3, whilst schools in arm 2 will deliver TWL to pupils in Year 4. In this ‘interleaved’ design, although randomisation will be at the school-level, all schools will receive the TWL programme in different interleaved year groups. The random allocation will determine which year group in the school receives the TWL programme. Pupils in year groups not receiving the programme in each arm will serve as controls for intervention pupils in the other arm. Note this design concept could be extended to more year groups with alternate year groups receiving the intervention interleaved with those who act as controls. There is a potential risk of contamination with this design as the same teacher or TA may teach both Year 3 and Year 4 pupils in a school. However, the focus of the intervention is activities delivered via the TWL platform which will not be available to control pupils, thereby minimising this risk.

The programme will be delivered to targeted pupils, using FFT’s Reading Assessment Programme (RAP) to select eight pupils per year group who will be part of the trial. Typically, these will be the lowest-performing eight pupils per year group although we note that pupil selection will be based on a combination of RAP scores and teachers’ professional judgement. Interested schools will complete an Expression of Interest (EoI) and schools identified as eligible will sign a Memorandum of Understanding (MoU) to confirm their participation. Half of the pupils in the participating year-groups will complete the baseline assessments, and the eight target children will be selected, prior to randomisation.

The primary outcome for this trial is reading attainment, which will be measured using the Woodcock Reading Mastery Test 3rd Edition (WRMT-III (Woodcock, 2011)). There are two sets of secondary outcomes for this trial. The first will use sub-tests within WRMT-III to measure oral reading fluency, reading comprehension and decoding separately. The second will use a paper-based survey with Year 4 pupils to assess children’s attitudes towards reading. This will use the eight items from section R7 of the student survey from Progress in International Reading Literacy Study (PIRLS) (Mullis and Martin, 2019; IEA, 2020).

Participant selection

School eligibility

The school-level eligibility criteria for schools are as follows:

- All two-form entry and larger schools with primary-aged children in England are eligible.
- Schools are not taking part in any other EEF funded trials within the same academic year.
- Schools are not offering any of the following programmes:
 - Accelerated Reader
 - Reciprocal Reading

- Dyslexia Gold
 - Lexia
 - ReadingPlus
 - Fresh Start (RWI)
 - Sparx Reader
- **Schools are not** participating in the follow Evidence into Action (EiA) projects: EiA in Worcestershire - Leading Literacy, EiA in Dudley - Reading Comprehension, EiA in Southwark - Disciplinary Reading or Havering - Scaffolding in reading.
 - Schools have not used TWL in 2023/24 or 2024/25
 - A maximum of ten schools from the same MAT can be recruited provided the individual schools meet all other eligibility criteria.

Pupil eligibility

TWL is a targeted intervention, designed to improve the reading skills of pupils achieving below age-related expectations. To identify pupils who are eligible to participate in this trial, we will follow a two-step process. In step 1, schools will nominate around 50% of pupils from each year group to be screened using RAP. Teachers will use their professional judgement to identify pupils for nomination, noting that the TWL programme cannot be accessed by pupils who are severely visually impaired or those who are mute. This pupil nomination step will take place in the summer term of 2025 when the pupils are in Year 2 and Year 3. In step 2 (September – October 2025), the nominated pupils, who will be in Year 3 and Year 4, will be screened using RAP. As there can be movement of pupils following the summer term (pupils leaving the school, new pupils joining), schools will have the opportunity to confirm the final list of nominated pupils in September 2025, prior to RAP testing. Prior to testing, FFT will conduct an online training session to provide guidance to teachers on administering RAP and using the RAP scores to select pupils to the trial. Teachers will use a combination of RAP scores and their professional judgement to select the final 8 pupils from each year group who will participate in the trial. It is our expectation that not all pupils who are nominated for screening will be eligible to receive the TWL programme based on their RAP scores. This is why we are asking schools to nominate more pupils than the final 8 who will be selected to the trial. Furthermore, pupils selected to receive the TWL programme will complete an initial assessment on the platform to determine their starting level and to pair pupils of similar abilities. Starting with a larger pool of pupils at the nomination stage will also give teachers the flexibility to select pupils that they judge to be of roughly similar levels to be selected to the trial to increase the opportunities for appropriate pairing within the programme.

Outcome measures

Primary outcome

We will use the Woodcock Reading Mastery Tests 3rd edition (WRMT-III; (Woodcock, 2011)) to assess the primary outcome. WRMT-III can be used to assess the reading skills of pupils between

the ages of 4.5 and 79 years. It consists of nine sub-tests that can be evaluated individually or combined into four cluster scores. The Total Reading Cluster score is calculated as the sum of the standardised scores of five sub-tests - Word Identification, Word Attack, Word Comprehension, Passage Comprehension and Oral Reading Fluency. For this evaluation, we will not use the Total Reading Cluster score based on the five sub-tests as the outcome measure. Instead, we will use four of these sub-tests that are aligned to the TWL intervention and of greatest relevance to us from the perspective of reading attainment. These are: Word Identification, Word Attack, Passage Comprehension and Oral Reading Fluency. As in the Lexia effectiveness trial (Merewood *et al.*, 2022), we will omit the Word Comprehension sub-test as the skills assessed by this sub-test repeat most of the skills assessed by the Passage Comprehension sub-test and including it would provide little additional information but add testing burden. We will use Principal Component Analysis (PCA), a dimension reduction technique, to create a composite score from these individual sub-test scores and this will be our primary outcome measure. To create this composite score, we will use the raw scores of pupils in each sub-test. Using raw scores rather than standardised scores is particularly important as we will combine the scores of pupils across Year 3 and Year 4. PCA will be used to derive the primary outcome measure for a number of reasons. Firstly, whilst the four subtests are designed to assess different aspects of reading ability, they are related and are likely to be highly correlated. PCA will help to identify the most informative combination of these subtests, reducing redundancy in the data. Secondly, PCA produces a composite score that weights each subtest based on its contribution to the overall variance. Finally, using the first principal component as the derived score ensures that maximum variation in reading ability across children is captured.

When these sub-tests are typically administered, the starting point (i.e. starting question or item) on each sub-test is based on the pupil's year group. For example, on the Word Identification sub-test, Year 3 pupils would start the test on item 12 whereas Year 4 pupils would start on item 16 of the sub-test. Since this trial will assess the impact of TWL on reading skills of Year 3 *and* 4 pupils, the scores of all pupils on all sub-tests will need to be placed on the same scale. Therefore, when we administer these sub-tests to pupils in the trial, we propose to use the starting point for Year 3 pupils as the starting point for both Year 3 and Year 4 pupils. This will ensure that all pupils are exposed to the same content on each sub-test and will allow us to combine the scores of pupils across both year groups. Since the pupils in this trial are performing/achieving below the expected standard in reading, we do not expect to see floor effects by taking this approach. At endpoint, the WRMT-III sub-tests will be administered one-to-one with pupils using paper-based assessments by trained NFER Test Administrators.

The WRMT-III was selected as the primary outcome measure because of its psychometric properties and comparability with the previous RCT evaluation of TWL. WRMT-III has a mean test-retest reliability of 0.92 by age. It has demonstrated correlation with other tests of similar outcomes such as Wechsler Individual Achievement Test 3rd edition (WIAT-III), Kaufmann Test of Educational Achievement, 2nd edition (KTEA-II) and the Woodcock Johnson III Tests of Achievement (WJ-III), which attest to the WRMT-III's convergent validity. In addition to these psychometric properties, the previous RCT evaluation of TWL used the some of the same sub-

tests (Word Identification, Word Attack, Passage Comprehension) to assess impact, which will facilitate comparability of some findings across the two trials.

Secondary outcomes

Secondary research questions will explore the impact of TWL on specific subgroups, on specific reading competencies, and attitudes towards reading (only Year 4 pupils). To answer these research questions, there will be two sets of secondary outcome measures. Firstly, subtests from the WRMT-III will be used to measure oral reading fluency (Oral Reading Fluency sub-test), reading comprehension (Passage Comprehension sub-test) and decoding (Word Identification and Word Attack sub-tests).

Secondly, attitudes towards learning will be measured via a paper-based survey that will use the eight items from section R7 of the student survey from Progress in International Reading Literacy Study (PIRLS) (IEA, 2020). This is designed to be age appropriate and accessible to pupils of this age group. The eight items are presented in Appendix A. Each question will be rated on a five-point Likert scale, with (perceived) negative responses scoring lower than (perceived) positive responses. For each pupil, a total score will be calculated by a simple summation of the individual item responses and will range between 8 and 40.

Baseline measures

For the primary outcome measure, the online reading diagnostic tool, the Reading Assessment Programme (RAP), that comes with the TWL programme and can measure reading fluency (Words Correct Per Minute), will be used to measure baseline reading skills. The raw score will be used as a baseline measure for the primary outcome and will additionally form the baseline measure for the secondary outcomes assessing pupils' reading competencies based on individual subtests of the WRMT-III. The RAP was selected as the baseline measure to minimise the burden of additional data collection since RAP is used in usual implementation of the TWL programme to identify suitable pupils. This also provides the added advantage of reducing testing costs for the trial. At baseline, the RAP will be administered one-to-one by teachers in schools.

The PIRLS-based survey administered to Year 4 pupils will also be used to collect baseline attitudes towards reading.

Sample size

Table 2: Sample size calculations

		All Targeted Children	FSM Children
Minimum Detectable Effect Size (MDES)*		0.180 (0.187)	0.233 (0.246)
Pre-test/ post-test correlations	level 1 (pupil)	0.662	0.662
	level 2 (class)	-	-
	level 3 (year group)	0.0	0.0
	level 4 (school)	0.600	0.600

Intraclass correlations (ICCs)**	level 2 (class)	-	-
	level 3 (year group)	0.0767	0.0767
	level 4 (school)	0.0383	0.0383
Alpha		0.05	0.05
Power		0.8	0.8
One-sided or two-sided?		Two-sided	Two-sided
Average cluster size (year group)		8	3
Number of schools	Intervention	72	72
	Control	72	72
	Total	72	72
Number of pupils	Intervention	576 (490)	216 (184)
	Control	576 (490)	216 (184)
	Total	1,152 (980)	432 (367)

* MDES estimates and numbers of pupils shown in brackets take into account 15% pupil-level attrition between randomisation and endline assessment.

** These sample size calculations are based on a three-level model that does not account for clustering at the class level. Although we expect class to contribute significantly to variation, the precise distribution of the ICC between class and year group is hard to estimate. For the actual analysis, we will use a four-level model.

A number of reviews provide expected effect sizes for EdTech programmes. The EEF 2019 evidence review on digital technology (Lewin *et al.*, 2019) finds effect sizes ranging between 0.16 to 0.22 for literacy and language based EdTech programmes. By contrast, a meta-analysis on the effectiveness of education technology on reading achievement found an average ES of 0.16 (Cheung and Slavin, 2012). Both suggest that our anticipated effect size of 0.18 for all targeted pupils is in line with results from other EdTech trials.

The assumed parameters for the sample size calculations are set out in Table 2 above. The calculations assume eight pupils per year group, with two year groups per school and 72 schools in each of the intervention and control groups. The assumed proportion of targeted pupils who have ever been eligible for FSM is based on the latest data from DfE (DfE, 2024), whilst ICCs and pre-post correlations have been taken from the latest EEF guidance on power calculations (Singh *et al.*, 2023), taking the median values based on NPD data between 2012 and 2019. The school and pupil numbers set out are those anticipated at the point of randomisation. The 72 schools we intend to recruit are higher than necessary to power the analysis sufficiently to account for any potential school-level attrition. However, we have not included any school-level attrition in these sample size calculations. The MDES and pupil numbers presented are for both no pupil attrition, and, in brackets, with 15% pupil-level attrition, which is in line with our experience in similar trials. We have assumed that total ICC will be split in the ratio 1/3:2/3 between school and year group, reflecting our assumption that the majority of variance would be accounted for at the

year-group level. To test this assumption, and its impact on the MDES, a series of sensitivity analyses were run that held all other parameters constant and varied the split of ICC across the two levels. The results of these sensitivity analyses can be found in Appendix C.

These sample size calculations are based on a three-level model that does not account for clustering at the class level, as reflected in Table 2 above. Although we expect class—more specifically, the class teacher—to contribute significantly to variation, the precise distribution of the ICC between class and year group is hard to estimate as most previous analyses use two-level models. For the actual analysis, we will use a four-level model, which more accurately represents the data structure and allows for direct estimation of ICC at each level. Additionally, the pre-test/post-test correlation at the year-group level (Level 3) is assumed to be zero, as its reliable estimation is hindered by the limited availability of studies providing sufficient data at this level. We note that setting the level 3 pre-test/post-test correlation to zero results in a more conservative MDES and, therefore, a more conservative sample size.

Sample size has been estimated using the PowerUpR (Bulus *et al.*, 2021) package within R. The syntax for the power calculations can be found in Appendix B.

Randomisation

Schools will be randomised into two arms in a 1:1 ratio.

- In arm 1, the 8 selected pupils in Year 3 will receive the TWL programme and the selected Year 4 pupils will not receive the programme.
- In arm 2, this will be reversed. The 8 selected pupils in Years 4 will receive the TWL programme and the selected Year 3 pupils will not receive the programme.

Pupils not receiving the programme in each arm will serve as the comparison group for pupils receiving the programme in the other arm.

Randomisation will be carried out for all settings at the same time (i.e. there will be no batches), and there will be no stratification when randomising. Randomisation will be performed by NFER statisticians using R Code, which will be stored for reproducibility and transparency. The syntax used for randomisation will be appended to the Statistical Analysis Plan (SAP) and report. Test administrators will be blind to group allocation, but analysts will not.

Statistical analysis

Primary and secondary analyses

Both primary and secondary analyses for this trial will be intention-to-treat and will follow the EEF analysis guidance (EEF, 2022b). The primary outcome measure is the first component from the PCA as outlined above and will be the dependent variable in a multi-level regression model with an independent indicator variable representing the status of the intervention for that student (1 = intervention, 0 = control).

The model will include four levels of clustering: pupil, class, year-group and school. The raw RAP score as outlined above will be used as the baseline measure and will be used as a covariate in this multi-level model that models both the class-level and the school-level clustering as random effects.

Secondary analyses will be carried out to assess RQ4 to RQ8. Analyses will be performed to assess the impact of TWL on specific reading competencies using WRMT-III subtest as outlined in RQ4 – RQ6 above. In each case, the analysis will use a multi-level model as outlined for the primary analysis, replacing the primary outcome measure with the respective subtest score and using the RAP raw score as the baseline covariate.

RQ8 will be assessed using a multi-level model, with a score derived from the PIRLS R7 survey responses as both outcome measure and baseline covariate.

All analysis will be carried out within R, with modelling performed using the nlme and lme4 packages.

Estimation of effect sizes

All effect sizes will be calculated using Hedge's *g*, using total variance from a model without covariates. 95% confidence intervals will be computed for the effect size.

Sub-group analyses

RQ2 and RQ3 will seek to analyse the impact of TWL on various subgroups:

- a. Disadvantaged children (FSM) (RQ2a). In alignment with EEF's focus on improving the educational attainment of pupils from disadvantaged backgrounds, we will assess the impact of TWL on FSM-eligible pupils' reading skills. A binary indicator that shows whether a child has been eligible for free school meals within the last six years (EVERFSM_6_P from schools via FFT) will be used as the identifier for disadvantage.
- b. Children with EAL (RQ2b) (obtained from schools via FFT)
- c. Children with SEN (RQ2c) (defined by converting the 'SENprovisionMajor' variable gathered from schools via FFT into a binary indicator (SEN versus no SEN))
- d. Individual year groups. For this analysis, data for each year group will be analysed separately, and the results from the two analyses will be compared. If effect sizes are broadly comparable (i.e. their confidence intervals overlap), this will indicate that there is no differential impact depending on year group.

The analyses for RQ2a, RQ2b and RQ2c will use a multi-level model similar to that outlined for the primary analysis. Each of these subgroup analyses will carry out two separate analyses. The first will assess the effect of the intervention just on that subgroup. The second will run a similar model to the one implemented for the primary analysis, with the addition of the subgroup indicator and an interaction term between the subgroup and treatment allocation.

Analysis in the presence of non-compliance

Compliance with TWL will focus on pupils' exposure to the programme and will assess whether pupils have actively participated in TWL as intended. Compliance will be measured at the pupil level using a continuous measure based on the number of weeks of compliant participation. Pupils who have completed 16 (or more) weeks of the programme attending at least 3 school-based sessions each week will be counted as 100% compliant.. Pupils with less than 16 weeks where they have attended at least 3 school-based sessions will be assigned a compliance percentage in proportion to the number of weeks they have participated in the programme. Note that while FFT recommends that schools deliver at least four sessions per week, the lower threshold of three sessions per week was chosen for compliance analysis based on FFT's expectation that a minimum of three sessions per week would be required to produce an impact for pupils. This brings the total number of sessions delivered to 48 over a 16-week period, closer to the lower end of the range of 50 – 60 sessions per pupil recommended by FFT.

We will explore the linear effect of this compliance indicator on the primary outcome measure. A complier average causal effect (CACE) estimate will be obtained using instrumental variable modelling (details will be provided in the SAP) and will be obtained for the primary outcome measure if the finding for the ITT analysis for the primary outcome measure is not statistically significant.

Additional analyses and robustness checks

Alongside analysing the effect of compliance on attainment, the impact of dosage of TWL on pupils' reading skills will be explored to assess how time spent on the platform (measured as the number of sessions attended) affects the impact of the programme. The number of TWL sessions at school and the number of sessions accessed at home will be analysed separately within one model. In both cases, dosage will be defined as a continuous numerical measure and will use the number of sessions of either activity for each pupil.

A moderator analysis will be carried out to examine the moderating effect of the change in reading enjoyment on pupils' reading attainment. The change in reading enjoyment will be measured using the attitudes towards reading data collected as the secondary outcome measure, with the primary outcome measure used to measure reading attainment. Details of these analyses will be provided in the SAP.

Missing data analysis

There is likely to be some degree of missing data for the primary outcome at endline. Where pupils are unavailable for testing, the reason for this will be established where possible and described in the final report. The complete case analysis will be completed if the volume of missing primary outcome at endline is less than 5%. If it is greater than 5%, we will explore the mechanism for missingness with additional analyses/checks. Additionally, variables that are found predictive of missingness in the primary outcome will be added to the primary analysis model to explore their impact on the estimated effect of the intervention. If required, further missing data analysis will

be undertaken if needed as per the EEF analysis guidance. Full analysis will be pre-specified in a detailed SAP.

Implementation and process evaluation (IPE) design

This IPE approach was confirmed as part of the set-up phase following the IDEA workshop, which provided the evaluation team with the opportunity to discuss the nature and theory of the TWL programme in-depth with FFT. This informed the evaluation team's understanding of the key elements of interest for the IPE, which has been designed to support understanding of the impact findings as well as how the programme could potentially be improved in future.

As there has only been one light-touch IPE conducted as part of a small RCT in the US (Ross, Laurenzano and Madden, 2017), this IPE has been designed with a focus on investigating the intended logic and mechanisms of the programme as outlined in the ToC, while further exploring engagement, outcomes and challenges to expand upon the findings from the earlier IPE. The roles and interactions of the EdTech platform, peer learning and Tutor support, which are intended to be complementary and mutually reinforcing, will be explored. The extent to which they are separately and/or in combination seen to impact outcomes will likewise be a point of focus. The implications of the home learning component of the programme in relation to equality of opportunity for pupils from disadvantaged backgrounds and/or those who may have relatively limited access to digital resources and support at home will be explored. Given the flexibility of the delivery schedule, we will assess the extent of variation in session timing and frequency to complement the dosage analysis. Finally, we will investigate potential moderators and relevant contextual factors, including year group, school resource and Senior Leadership Team (SLT) support.

Research questions

Fidelity & adaptation

IPE RQ 1: *To what extent was the TWL programme delivered as intended?*

- Did the nominated SCs and Tutors complete the training and to what extent did it prepare them for programme delivery?
- What criteria and process did schools follow in selecting and pairing pupils for the programme, and to what extent did this align with the guidance provided by FFT?
- What was the nature and extent of:
 - FFT support for schools?
 - SC support for Tutors?
 - Tutor support for pupils during the sessions, including monitoring of progress and adjusting levels of support accordingly?

- What was the level and nature of variation in how many sessions were delivered and when, as well as the number of story levels completed?
- To what extent did home use of the programme and the fostering of home-school links occur, and what did this look like?
- What adaptations to the programme were observed and what were the reasons for them?

This research question will explore the extent to which the programme was delivered as intended. The findings will provide important context for interpreting the results of the impact analysis in terms of understanding whether they reflect the outcomes of the programme's theory in practice or whether significant variation in implementation occurred. These IPE findings will be complemented by the dosage analysis that will look specifically at the amount of programme received. We are particularly interested in understanding what the roles of the Tutor and SC look like in practice given the training and preparation they receive, and the extent to which this varies across the sample. As the Independent Access element of the programme was introduced shortly before the trial, the extent to which its use is encouraged by schools and taken by pupils will be important to observe. In addition to qualitative data collected through methods such as interviews and observations, we will use the TWL platform analytics data to provide quantitative indicators of fidelity and carry out regression models to understand their effects (see 'Analysis' below). Finally, the nature and extent of adaptations made to the programme will inform ongoing considerations around how the programme may be improved to best meet school needs and to encourage fidelity of implementation.

Context & moderators

IPE RQ 2: *What are the key moderators and contextual factors that influenced how effectively the programme was delivered?*

- What were the key challenges and facilitators for successful implementation of the programme, including for specific pupil sub-groups (FSM, EAL, SEND, different year groups)?
- To what extent were levels of parental engagement, home environment, home use of the programme and fostering of school-home links seen to moderate impact?
- To what extent were pupil attitudes to reading seen to moderate pupil reading outcomes?

This research question is concerned with understanding how implementation of the programme may have varied depending on the context in which it was delivered and/or the characteristics of the staff and pupils involved. School resource, space and support from the SLT are likely to be key moderators in this respect, as are pupil and Tutor competencies with using EdTech tools. Identifying challenges for pupils eligible for FSM, with EAL and/or with SEND will help us to better understand the findings from the impact sub-group analysis (RQ2) in terms of whether the programme is more or less effective for them and/or whether they may face additional challenges in terms of benefitting from it. The role of the home environment on the extent to which a child is able to benefit from the programme will be explored. Whether changes in pupil attitudes towards reading may moderate any change in terms of reading outcomes will be qualitatively examined to

complement the moderator analysis (see ‘Impact evaluation design’). A number of other relevant contextual factors and/or moderators may emerge from our data collection and a more exhaustive list of those we anticipate observing is provided in the logic model.

Pupil responsiveness

IPE RQ 3: *To what extent did pupils engage with the programme, including peer learning opportunities?*

- To what extent did pupils engage with the co-operative learning element of the programme, and what did this look like in practice?
- How effective was the programme at engaging and meeting the needs of the target group, including in terms of platform content and interface?
- To what extent did the programme engage and meet the needs of specific pupil sub-groups (FSM, EAL, SEND, different year groups)?

The nature of the TWL programme requires high levels of pupil engagement due to use of the platform being self-directed and co-operative learning being a key theorised mechanism of change. As a result, we are particularly interested in understanding how the co-operative learning element plays out in practice and to what extent it appears to vary between children and between schools, and which (if any) factors may be driving this. Pupil enjoyment of the sessions and motivation to work on the platform will be qualitatively gauged to help inform our interpretation of any change in reading attitudes that may be observed in the impact analysis (RQ8). The potential effect of task and content repetition, which is part of the programme design, will also be considered in relation to pupil engagement and motivation.

Perceived impact

IPE RQ 4: *What was the perceived impact of the programme for (i) teachers/TAs, (ii) pupils and (iii) disadvantaged pupils specifically?*

- To what extent was the programme perceived to have an impact on teacher/TA professional skills and subject knowledge in relation to supporting reading development (through phonics)?
- To what extent was the programme perceived to have an impact for *all* pupils and for disadvantaged pupils specifically in relation to:
 - Pupil attitudes towards reading – including enjoyment and confidence of reading (also covered by the IE)?
 - Cooperative learning behaviour and skills?
 - Levels of home support and engagement around reading?
- What (if any) unintended impacts were observed – whether positive or negative?

While the impact analysis will provide an estimate of programme impact on pre-specified primary and secondary pupil outcomes, this research question will focus on understanding a broader range of possible benefits as perceived by the participants (both pupils and staff). We are particularly interested in understanding any perceived impacts for the teachers/TAs or the pupils' home environments, given the presence of these threads through the logic model and the limited evidence regarding these elements thus far. We will also look for any unintended impacts that were observed, which may include negative ones such as learning displacement due to the pupils missing lesson time to do the programme, or indeed for pupils *not* participating in the programme who have missed out on the teacher/TA support they would usually receive as a result.

Mediators

IPE RQ 5: *What were the perceived mediators for any pupil outcomes observed?*

- To what extent were teacher/TA outcomes seen to be a mediator for pupil attainment outcomes?
- What was the perceived value of specific programme elements (e.g. online training, EdTech, co-operative learning, Tutor support)?

Several possible mediating pathways are outlined in the logic model that will not be within the scope of the impact evaluation to analyse. As a result, this research question will focus on understanding perceptions of these possible mediating pathways for further investigation in future. The role of teacher/TA outcomes in determining pupil outcomes will be explored to understand both the relative importance of the Tutor support provided and the potential value of the programme as broader continuous professional development (CPD). We will also look to understand the extent to which any impact on pupil reading outcomes is driven by changes in pupils' attitudes to reading and/or in their home learning environment. If a programme impact is observed, this question will also help to understand the extent to which specific programme elements (e.g. the platform, co-operative learning, Tutor support) may be driving it and/or whether the complementary nature of these elements is significant.

Programme differentiation

IPE RQ 6: *What was business as usual (BAU) and to what extent did it differ from the TWL programme?*

- What was the nature of BAU in all schools prior to the programme implementation period?
- What was the nature of BAU in control schools during the programme implementation period?
- To what extent did the programme provide different opportunities for participating pupils compared to BAU?
- What (if any) contamination occurred via participating teacher/TA support for control year groups?

This question will complement the impact analysis by seeking to understand the extent to which practice in control and treatment schools differed and, hence, the extent to which any difference in outcomes may be attributable to the programme, or indeed a likely underestimate of the true effect. We will also use this question to examine the extent to which the core components of the programme (as outlined in the 'Intervention' section and elaborated in the ToC) represent a genuine departure from what schools already had in place. For example, even if the EdTech component of the programme is new, participating pupils may already have been doing regular small group reading development sessions with their teacher/TA using similar techniques. Finally, given the interleaved design of this trial, this question will look to understand the extent to which any contamination occurred between the year groups designated as treatment and control. In particular, we will look to understand whether the participating teachers/TAs has provided any support to pupils in the control year group since the start of programme delivery and/or carried out any significant knowledge sharing activities with colleagues working with those pupils.

Research methods

The IPE employs a range of mixed-methods data collection activities to provide data from different perspectives that can be triangulated to form a richer picture of the process of implementation. The IPE data collection activities occur in a staggered manner over the course of the trial. The activities that will occur in each phase of the trial are listed in Table 3. This enables each data collection instrument to be informed by the context and emerging findings from the previous activity. For example, the training observations and review will help to build a strong understanding of the programme that can help to assess fidelity during the case study observations. Similarly, data collected from the case study activities will be important for identifying the key points of interest for the endpoint surveys and in formulating how best to address them.

All data collection instruments will be informed by the programme logic model, TIDieR framework (see the 'Intervention' section) and IPE research questions, in addition to findings drawn from the data already collected.

Table 3 Sequencing of IPE data collection activities

Pre-randomisation	During training	During delivery	Post-delivery
SC survey (all)	Semi-structured observation of Tutor training	Case study observations, interviews & pupil focus groups (x8)	SC survey (all)
			Tutor survey (treatment)
Semi-structured observation of RAP training	Semi-structured review of support materials		Pupil Survey (all) ⁷
	Training completion data		TWL platform analytics

⁷ The endpoint pupil survey will collect data for both the impact evaluation and IPE. The impact questions of the survey will be administered to all pupils prior to randomisation as well.

Table 4: IPE methods overview

IPE dimension	RQ	Research methods	Data collection methods	Sample size and sampling criteria	Data analysis methods
Fidelity & adaptation	1	Qualitative methods	Semi-structured observations	1 x RAP training session, 1 x Tutor training day, 8 x programme sessions (across 8 schools)	Thematic analysis
			Semi-structured review	Training & support materials	
			Semi-structured interviews	8 x SCs, 8 x Tutors (across 8 treatment year groups)	
			Focus groups	8 x groups of 4 pupils (across 8 treatment year groups)	
		Quantitative methods	Surveys	SC survey (all, endpoint), Tutor survey (intervention, endpoint), pupil survey (all, endpoint)	Descriptive statistics
			Monitoring data	TWL platform analytics, training attendance data	
Context & moderators	2	Qualitative methods	Semi-structured observations	8 x programme sessions (across 8 schools)	Thematic analysis
			Semi-structured interviews	8 x SCs, 8 x Tutors (across 8 treatment year groups)	
			Focus groups	8 x groups of 4 pupils (across 8 treatment year groups)	
		Quantitative methods	Surveys	SC survey (all, endpoint), Tutor survey (intervention, endpoint), pupil survey (all, endpoint)	Descriptive statistics, linear multi-level modelling

IPE dimension	RQ	Research methods	Data collection methods	Sample size and sampling criteria	Data analysis methods
			Monitoring data	TWL platform analytics, training attendance data	
Pupil responsiveness	3	Qualitative methods	Semi-structured observations	8 x programme sessions (across 8 schools)	Thematic analysis
			Semi-structured interviews	8 x SCs, 8 x Tutors (across 8 treatment year groups)	
			Focus groups	8 x groups of 4 pupils (across 8 treatment year groups)	
		Quantitative methods	Surveys	SC survey (all, endpoint), Tutor survey (intervention, endpoint), pupil survey (all, endpoint)	Descriptive statistics
			Monitoring data	TWL platform analytics, training attendance data	
Perceived impact	4	Qualitative methods	Semi-structured interviews	8 x SCs, 8 x Tutors (across 8 treatment year groups)	Thematic analysis
			Focus groups	8 x groups of 4 pupils (across 8 treatment year groups)	
		Quantitative methods	Surveys	SC survey (all, endpoint), Tutor survey (intervention, endpoint), pupil survey (all, endpoint)	Descriptive statistics, linear multi-level modelling

IPE dimension	RQ	Research methods	Data collection methods	Sample size and sampling criteria	Data analysis methods
Mediators	5	Qualitative methods	Semi-structured interviews	8 x SCs, 8 x Tutors (across 8 treatment year groups)	Thematic analysis
		Quantitative methods	Surveys	SC survey (all, endpoint), Tutor survey (intervention, endpoint)	Descriptive statistics
Programme differentiation	6	Qualitative methods	Semi-structured interviews	8 x SCs, 8 x Tutors (across 8 treatment year groups)	Thematic analysis
		Quantitative methods	Surveys	SC survey (all, baseline & endpoint), Tutor survey (intervention, endpoint)	Descriptive statistics

1. Reviews and observations of training and support material

The nature and extent of the training and support provided to schools will be understood through semi-structured observations of the live sessions and reviews of the support material provided. This will inform the development of the subsequent data collection instruments and provide important context for later data collection activities, particularly for determining fidelity during the case study observations. Feedback from SCs and Tutors regarding their experience of the training and support will also be triangulated with this data to provide a more nuanced picture of the programme's strengths and where it could be improved.

Semi-structured observations of the RAP training and Tutor training

An NFER researcher will attend and observe one of the online RAP training sessions and one of the online Tutor training sessions. The observations will be semi-structured, with the researcher noting down points of interest in relation to the training structure, level of interaction and trainee engagement, and the training content itself. There will be an observation tool to guide the data collection based on the TIDieR framework and logic model.

Semi-structured review of support material

The live training observations will be supplemented by a semi-structured review of the support materials provided for schools, including the Support Guide for Tutors. As with the observations, an NFER researcher will read the materials, noting down key points on the structure and kinds of information provided, guided by a tool based on the TIDieR framework and logic model. This review will be completed prior to the start of programme delivery and will help to strengthen the team's understanding of the programme, allowing for greater nuance in later explorations of fidelity.

2. Case studies

Case studies will be carried out with eight schools over the course of the delivery period to help us develop an in-depth understanding of various IPE dimensions. While this number of case study schools means that the sample cannot be seen as representative, it will still allow for variety in contextual factors that are likely to affect implementation – particularly year group and school-level proportion of FSM. We will also look to use the TWL platform analytics data to sample schools across a range of engagement levels. At the same time, only a small number of schools will be affected by the additional research burden.

Based on the criteria outlined above, we will invite specific schools to accept or decline an invitation to participate in the case study data collection. We will clearly communicate what being a case study school involves and what will be expected of participating staff members before confirming the school as a case study. The majority of the data collection activities will take place on a single visit to the school, although remote follow-up interviews may occur when convenient to the school. We will work closely with the key contact at the school to confirm the timings of each data collection activity in advance of the visit.

Case study visits will be undertaken from week 3 of project delivery onwards, with the aim of spacing data collection out to ensure a range of delivery stages are covered by the case studies overall.

Case study schools will receive a ‘thank you’ book voucher of £150 in recognition of the additional time they give to the evaluation, as well as to encourage participation from schools regardless of their level of engagement in the programme itself.

Triangulation of the observation data and feedback from two staff members and four participating pupils will allow us to build a strong, nuanced picture of implementation in that school, any challenges and/or adaptations involved, the impact it is perceived to be having and the extent to which it is being delivered with fidelity.

Semi-structured observations of session delivery

One observation of a session taking place will be carried out per case study. During the case study visit, the NFER researcher will sit in the room with the children and Tutor while one of the sessions is delivered. The researcher will aim to be as unobtrusive as possible so as to minimise impact on the dynamics of the session. When observing the session, the researcher will focus on the level and nature of support provided by the Tutor, levels of pupil engagement (particularly among disadvantaged pupils)⁸ and any relevant contextual factors, such as the space in which the session is delivered. As each session is guided by the programme, the observation will be more focused on the fidelity of the set-up rather than the specific content covered in the session. This includes whether there is sufficient space for the two pairs of pupils to work without disturbing each other, whether the technology and internet available is of sufficient quality and how long the session lasts.

In addition, one pair of pupils⁹ will be video-recorded during the session to allow for a more in-depth study of the nature and extent of co-operative learning practices occurring during the sessions without intruding on the natural dynamics between the two pupils through close in-person observation. This data collection activity will help us to understand the nature and level of pupil responsiveness to the programme. The video recording will ideally be achieved through the computers that the pupils are working on via a Teams call set up from the researcher’s own laptop. Where this is not possible, the researcher will use the camera on their laptop to record the pupils’ interactions from a non-obtrusive angle. The researcher will observe this video recording to take further notes at a later date.

⁸ The FSM eligibility status of all pupils within the intervention group for that school will be identified in advance of the visit through the data analytics.

⁹ If parental consent is obtained for all pupils in the session, then the video recorded pair will be selected pseudo-randomly. If parental consent for both pupils within a pair is only obtained for one of the pairs, they will be selected for the video recording. If parental consent is not obtained for both pupils in either pair then no video recording will occur.

Parents of all children involved in the session will be asked to provide explicit opt-in consent for both the video recording and pupil focus groups (see below). The pupils themselves will likewise be asked for consent for both of these elements following an age-appropriate explanation of what they involve.

The main observation tool will be designed to guide the researcher in noting down points of interest in relation to fidelity, adaptations, context, challenges and facilitators and pupil engagement. In addition, the researcher will use a framework for observing the presence of cooperative learning behaviours based on the items in Johnson & Johnson's student-student interaction checklist for teachers (Johnson and Johnson, 1975). This will be designed to elicit qualitative observations relating to whether the pupils are pair- or self-oriented in completing the tasks, how attentive the pupils are to each other and whether the pupils perceive the other as a source of assistance.

Semi-structured SC and Tutor interviews

Each case study will include an interview with the SC and participating Tutor. These interviews may occur together or separately, depending on the interviewees' preference and availability. The interviews can take place in-person following the observation on the day of the school visit, or they can take place remotely on a later date using Microsoft Teams. Each interview will last between 30 and 45 minutes. The interviews will be video recorded using Microsoft Teams if conducted remotely (audio-only if in person) for note-taking purposes.

The SC interview will focus on the more structural aspects of programme implementation, such as potential challenges relating to scheduling, space and material, the process of selecting and pairing the pupils, and any use of the Independent Access component. The Tutor interview will be centred on the sessions themselves, including their perception of the Tutor role, pupil engagement and adaptations made. Both interviews will cover support levels from FFT and their school SLT, perceived outcomes for the Tutor and pupils, and the perceived value of the programme training, support material, intervention model and the platform's data analytics. Given the interleaved trial design, we will ask about any contamination that may have occurred, for example through the Tutor providing support to the control year group. We will also ask for the interviewees' views on the extent to which the programme differs from their usual practice, and any perceived risks or drawbacks, such as learning displacement or reduced support elsewhere. Interviewee observations relating to disadvantaged pupils specifically will be drawn out as much as possible.

Pupil focus groups

Following the observation, a pupil focus group will be carried out with the four pupils involved in the session (excluding any pupils for whom opt-in consent had not been obtained from either the parents or the pupils themselves). These focus groups will last between 15 and 30 minutes. They will be age-appropriate and use engaging visual material. The pupils will be asked about what they do and do not like about the programme, with prompts such as working with their partner, the platform activities, using a computer and winning prizes. We will also ask them about how the programme makes them feel about reading. Finally, we will ask whether they read much at

home and how often their family member(s) reads with them. Where relevant, we will ask them about their use of the Independent Access component.

We are aware of the risk that questions about the pupils' home environment may result in a safeguarding disclosure. All researchers involved in these data collection activities will have up-to-date safeguarding training and follow the school and NFER's safeguarding protocols in all instances. A member of school staff will be present at all times during the focus group.

SC surveys

The nominated SC for all participating schools will complete a 10-15-minute online survey at both baseline and endpoint.¹⁰

At baseline, the survey will ask about existing practice in both Years 3 and 4 in relation to support for phonics and reading – both in the classroom and via specific interventions. The survey will also ask about use of Edtech and cooperative learning strategies within the school. At endpoint, these same questions will be asked in relation to the year group allocated to control over the intervention delivery period. This will help us to understand the extent to which programme implementation may have varied between schools depending on their learning tools and practices in the comparison year group.

In addition, the endpoint survey will ask questions about implementation of the programme in the year group allocated to the intervention. These questions will cover similar topics to the SC interviews, with the aim of understanding the prevalence of different experiences that may have been explored more in-depth through the case study data. The survey will also look to understand whether any direct contamination may have occurred through SC work with the year group allocated to control, or indirect contamination as a result of SC sharing learning with other staff members in the school.

Finally, the survey will cover questions relating to the time and costs incurred by schools to deliver the programme (see 'Cost evaluation').

Tutor survey

The lead participating Tutor per school will be asked to complete an online survey at endpoint about their experience of delivering the programme. As with the SC survey, this will cover similar questions to those in the Tutor interviews, with the aim of understanding the breadth of some of the observations that emerged from the case study data. The survey will take approximately 10 minutes to complete.

Pupil surveys

Paper-based surveys will be administered to all Year 4 pupils in the trial at baseline and endpoint to collect data for the impact analysis (see Impact section). Question R6 from the 2021 PIRLS

¹⁰ All online surveys will be administered using the Tivian survey platform.

Student Questionnaire on frequency of reading outside of school (IEA, 2020) will be adapted to focus on frequency of reading with family members as a proxy variable for the quality of home support for reading.

In addition, pupils allocated to the intervention will be asked close-ended questions about what they liked and did not like about the programme, how it made them feel about reading and how much (if at all) they accessed the platform at home.

As with the pupil focus groups, NFER's Child Protection Protocol will be followed in the unlikely event of any safeguarding disclosures occurring through the pupil survey. However, the closed-response nature of the questions will reduce the risk of this occurring.

3. Monitoring data

Training attendance data

FFT will share with NFER in anonymised form the number of attendees at the RAP and TWL training sessions per school to support understanding of fidelity.

TWL platform analytics

Pupil-level data analytics from the TWL platform will be shared by FFT for NFER to use to better understand fidelity, dosage and potentially moderating factors.

The programme is designed to be delivered in short but frequent sessions, resulting in the recommendation of at least four sessions per week. We will look at the median number of sessions per week for each child to understand levels of fidelity in this respect (i.e., how common it is for a pupil to complete fewer than three sessions per week).

While working in the Free Play setting is recommended for a pupil whose partner is absent for one or two sessions, it is not theorised to be as effective for supporting reading development as usual programme practice so should not be used longer term. We will use the data analytics to look at the number of Free Play sessions completed per child to understand the extent of variation in this respect and what implications this might have for the findings of the impact analysis. For example, a high prevalence of Free Play sessions may reduce the effect that would have been seen with higher rates of standard practice.

We will look at the prevalence of pupils working as individuals, rather than in a pair, for similar reasons. While the programme stipulates that it is appropriate for children to work independently if needed, it is not the desired implementation model so may influence the programme's impact on pupil outcomes.

Having pupils working at an appropriate story level for their attainment is a core element of the ToC and thus an important consideration for fidelity. The platform analytics offer several proxies for understanding this at the pupil level. For example, pupils are expected to complete an average of 0.3 stories per session (i.e. approximately three sessions per story). Pupils with a much lower or higher story completion rate are likely not to be working at the appropriate level, indicating poor fidelity and likely having implications for any impact observed.

Similarly, Tutor Check WCPM scores that are marked as 'out-of-range' indicate that the pupil achieved a much higher score than would be anticipated, suggesting that the pupil is not working at the appropriate story level for their attainment. As with the rate of story progression, we will look at the number of 'out-of-range' Tutor Checks per child to identify the prevalence of this kind of poor fidelity and reflect on potential implications for pupil outcomes. The difference between the story level allocated to the pupil by the TWL assessment and the pupil's actual starting story will likewise be used to understand whether pupil pairings may have resulted in pupils not necessarily working at the story level most suited to their reading attainment.

At the most basic level, we will look to understand dosage across the sample in terms of (a) time, through the total number of sessions completed, and (b) content, through the number of story levels completed. In addition, we will look at variation in the amount of Independent Access sessions completed.

Each of the above metrics will be broken down by the sub-groups of FSM eligibility, EAL, SEND and year group. This will enable us to see whether there are any clear differences in levels of fidelity and/or dosage for pupils within these sub-groups compared to their peers.

This data will be triangulated with the self-reported experiential data from the case studies and surveys to better understand what programme delivery looked like in practice.

Analysis

All IPE analysis will be carried out in line with the EEF's guidance (EEF, 2022a).

Qualitative data – interviews, focus groups, observations and reviews

Observation and review data will be treated qualitatively to provide a clear narrative of what the programme looks like in practice and the key variables that influence the nature of its implementation. Notes from the observations, reviews, interviews and focus groups will be written up as intelligent verbatim transcripts¹¹ and uploaded to the qualitative data analysis software MAXQDA (VERBI Software, 2021). The data will then be analysed using thematic analysis to identify and analyse both manifest and latent patterns of meaning (Joffe, 2012). This will involve an initial process of deductive coding from a pre-determined coding frame based on the ToC, IPE dimensions and research questions, as well as preliminary findings from the IPE data collection activities, followed by a more detailed inductive coding process to draw out the key themes emerging from the texts.

The MAXQDA 'variables' function will support within-case analysis by making it possible to look at all the data from a case study for each code in one place (MAXQDA, 2022). This will help us to triangulate the different data sources and better understand the context in which particular

¹¹ Intelligent verbatim transcription excludes fillers and redundancies that do not add meaning to the content to make the text more 'readable' (McCullin, 2023).

issues or perspectives emerge. At the same time, inter-case study analysis will be facilitated by the option to view all data per code across one or more data source types.

At least two researchers will be involved in the coding process to reduce the risk of bias. A common approach will be discussed and agreed upon by the coding team in advance, and an initial subset of the data will be double-coded and discussed to support inter-coder reliability. This will be facilitated by the use of MAXQDA's Intercoder Agreement function, which facilitates comparison of the independent coding of the same text performed by two different people (MAXQDA, 2025). Further cross-checking will occur whenever necessary, with questions and uncertainties addressed collaboratively by the coding team.

Quantitative data – surveys and administrative data

Survey response data will be exported from Tivian and quality assured prior to its analysis, with the data from each source stored in a separate file.

The analysis plan for the survey and platform analytics data will be designed by the team statistician in consultation with the IPE Lead, Project Lead and Project Director. The statistician will then carry out the analysis using R, with tables and charts produced in Microsoft Excel. All codes and outputs will be checked by another experienced member of NFER's Centre for Statistics. The analysis will cover a range of descriptive statistics as required to address the IPE research questions, including a small number of cross-tabulations of key variables to better understand the relationship between them. These variables will be selected in advance as part of the analysis plan. For example, we may look to understand if and how staff survey responses vary according to school-level disadvantage (through the proxy of the proportion of pupils eligible for FSM).

Pupil survey data will be matched with FSM eligibility to enable cross-tabulations of pupil responses by level of disadvantage (with FSM eligibility as the proxy).

In addition, we will use linear multi-level modelling to statistically test for differences in pupil-reported home environment outcomes between treatment and control groups at endpoint, controlling for baseline scores as a covariate.

Finally, we will use linear multi-level modelling to statistically test the degree to which various aspects of fidelity (Free Play sessions, whether pupil worked as a pair or not, number of out-of-range tutor checks, difference in levels between assessment story and starting story) affect outcomes, controlling for baseline scores as a covariate.

Triangulation of qualitative and quantitative data

The data collection tools across both the qualitative and quantitative methods will be designed to facilitate triangulation and complementary analysis. For example, certain questions will be asked in both the Tutor interviews and Tutor survey, to allow for both an in-depth understanding of that particular dimension as well as the prevalence of the various perspectives that emerge. Considerations relevant to future data collection tools that emerge from the data collection activities will be logged in real-time and referred to during the instrument design process.

We will also develop an integrated analysis framework to guide our analysis and reporting for each of the IPE dimensions and research questions. An IPE analysis workshop will be carried out prior to report writing to cross-check interpretations of the data and to develop an outline of the key points to be covered under each of the research questions. A subsequent analysis workshop with the broader evaluation team will bring the IPE and impact analysis findings together to facilitate reflection on any insights that the IPE may bring to the impacts that may or may not have been observed.

Cost evaluation design

There has not yet been any cost evaluation of the TWL programme conducted. As a result, we will carefully monitor the cost in terms of both time and expenditure that intervention schools incurred as a result of implementing the programme as part of the trial. We will collect information on the pre-requisite, set-up and ongoing costs to schools, in accordance with the EEF cost evaluation guidance (EEF, 2023).

To minimise data collection burden on schools we will collect cost evaluation data from FFT wherever possible. FFT will provide the training attendance data and information on the length of the RAP and TWL training sessions that will enable us to determine staff time spent on training. FFT will also provide information on any devices provided to schools for the purpose of delivery.

The main cost evaluation data collection tool will be the cost pro forma that will be administered to the IPE case study schools. The cost pro forma will ask about staff time spent on training and programme delivery and any additional costs, including staff cover, differentiated by role. The SC at each case study school will be provided with a digital copy of the cost pro forma in advance of the case study visit and will be asked to submit a completed version to the IPE Lead via email at the end of the delivery period. The case study visit will provide an opportunity for the SC to ask any questions they may have about completing it.

The cost pro forma will focus on eliciting detail about time and costs that it would not be possible to obtain through broader IPE instruments or FFT. However, the SC and Tutor surveys will also include low-burden cost questions that will provide context on how representative of the intervention schools the case study sample may be seen to be. This will minimise burden on schools while enabling triangulation from various sources.

We will ask schools about time spent:

- Completing the RAP and TWL training (including the cost of supply cover if training time was not covered internally)
- Preparing for programme delivery *in advance* of the first session
- Preparing for each week's session, including any additional top-up training (e.g. watching on-demand videos)
- Monitoring pupil progress and support needs

- Scheduling the programme sessions and securing necessary space and equipment
- Delivering the sessions
- On any other related activities

Estimates of time spent will be recorded separately for the Tutor and SC. Given that Tutor and SC roles are not tied to a particular pay grade, we will ask all participants to specify their role. Our main analysis will focus on the most common roles associated with these responsibilities. We will also conduct a sensitivity analysis accounting for the different pay grades that were observed. We will also look to understand the extent to which time spent on this programme was in addition to or instead of the TA/teacher's usual workload and, if the former, whether this additional time was paid.

We will collect data on the cost of any cover time that had to be arranged to enable teachers to spend time on the training and programme delivery, as well as any cases where additional staff members were recruited for this purpose. Wider staffing implications this may have had for the schools will be captured in descriptive terms.

We will look to understand any pre-requisite costs, with the key focus for this programme being the laptops/computers required for the sessions, as well as a stable internet connection. While these elements will not be included in the main cost estimates of the programmes, we will look to understand for how many schools these elements constituted an additional cost and carry out sensitivity analysis to account for instances when this material may or may not already be available. Any other unanticipated costs such as printing, stationary etc. will likewise be accounted for.

The main financial cost to schools is expected to be the cost of the programme. Given that the programme is already available for purchase we will be able to use the existing market price of the programme in our analysis. We will confirm this price directly with FFT.

The cost evaluation analysis will be carried out in line with the EEF's latest cost evaluation guidance. Each cost to the schools will be estimated per year over a projected three-year period. We will differentiate between pre-requisite costs, start-up costs and recurring costs. Ongoing costs in years 2 and 3 will either be reduced to zero (for fixed cost) or we will make an informed decision about whether they are likely to change over time. Time and cost estimates will be reported in terms of means and ranges. Once we have established the cost per-school-per year, we will divide this by the number of intervention pupils per school to estimate the cost per-pupil-per-year.

Ethics and registration

This evaluation will be conducted in accordance with NFER's Code of Practice, available at [NFER Code of Practice](#). All of NFER's projects abide by its Code of Practice, which is in line with the Codes of Practice from BERA (the British Educational Research Association), MRA (the Market Research Association) and SRA (the Social Research Association), among others. NFER is

committed to the highest ethical standards in all of its activities and ethical considerations are embedded in its detailed quality assurance processes.

This trial will be registered at the [ISRCTN](#) registry and the trial registration details will be provided in the Statistical Analysis Plan (SAP). The trial registry will also be updated with outcomes at the end of the project.

Each participating school's headteacher will provide their agreement to participate in the trial by signing the Memorandum of Understanding (MoU) that outlines the responsibilities of all parties involved in the trial. NFER will share a parent letter and withdrawal form with schools to be sent to parents/carers of all pupils that schools intend to nominate for participation in the trial. Through the withdrawal form, parents/carers will have the opportunity to withdraw their child from the evaluation and associated data processing at any stage of the trial. Unless explicitly instructed, NFER will use all data collected up to the point of withdrawal in analysis. Withdrawal of pupils' data will not be possible after analysis has commenced (August 2026).

A separate opt-in agreement process will be used for the observation and video recording of tutoring sessions and pupil focus groups and will only apply to those selected to participate. Since pupils participating in the tutoring sessions and focus groups are only 7 to 9-years-old, we cannot assume that all pupils will have the capacity to provide fully informed consent to participate. In addition, as the tutoring session observations and focus groups involve video and audio recordings respectively, it is especially important to ensure that parents/carers have the option to specifically agree to their child participating in these evaluation activities. We will, therefore, provide parents/carers with a written information sheet which will contain full details about the tutoring session observation and focus group and what their child will be asked to do. Parents/carers will then be asked to provide written opt-in agreement for their child to be invited to participate in the focus group, by returning a consent form to the school, who will then pass this information on to the research team.

Pupil participation in these activities is voluntary, therefore even if a parent/carer has agreed for their child to participate, their child can still choose not to take part. Age-appropriate information about the tutoring session observations and focus groups will be provided to pupils at the same time as parents/carers receive information about these activities to allow them to discuss participation together. NFER researchers will also read this information to pupils at the beginning of each of these activities to ensure pupils understand it and have the chance to ask any questions. If at any point a pupil decides that they would prefer not to participate, then they will be able to return to their class. Prior to beginning the focus group, the researchers will agree some ground rules for the group with the pupils and have a discussion with them about the types of scenarios in which we would need to break confidentiality, to ensure they fully understand what this means.

Data protection

Please see the [Data Protection Statement](#) for EEF Evaluations.

All data gathered during the trial will be held in accordance with the data protection framework created by the Data Protection Act 2018 and the General Data Protection Regulation 2016/679 and will be treated in the strictest confidence by the NFER, FFT and EEF. No individual or school will be identified in any report.

NFER and FFT will sign a Data Sharing Agreement that will govern the collection and sharing of personal data during this trial. This agreement includes a description of the nature of the data being collected and how it will be shared, stored, protected and reported by each party. In addition, FFT will provide a memorandum of understanding to schools, explaining the nature of the data being requested of schools, teachers and pupils, how it will be collected, and how it will be passed to and shared with NFER. Two separate Privacy Notices, one for schools and another one for parents, are available [here](#).

Legal Basis

The legal basis for processing **personal data** is covered by: GDPR Article 6 (1) (f) which states that ‘processing is necessary for the purposes of the **legitimate interests** pursued by the controller or by a third party except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of the personal data’. We have carried out a legitimate interest assessment, which demonstrates that the evaluation fulfils one of NFER’s core business purposes (undertaking research, evaluation and information activities). It also has broader societal benefits and will contribute to improving the lives of learners and teachers by providing evidence of the impact of online tutoring platforms on learning outcomes and classroom practice. Therefore, it is in our legitimate interest to process and analyse personal data for the administration of this RCT and the analysis of its impact on reading outcomes for pupils. Details of all data processed by NFER for this project are also recorded in the project’s data log that is overseen by NFER’s Compliance Officer.

The legal basis for processing pupils’ **special personal data** (e.g. SEN data) is covered by GDPR Article 9 (2) (j) which states that ‘processing is necessary for **archiving purposes in the public interest, scientific or historical research purposes or statistical purposes**’ in accordance with Article 89(1) (as supplemented by section 19 of the 2018 Act) based on domestic law which shall be proportionate to the aim pursued, respect the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject’.

Data Collection and Sharing

School staff Data

For the purposes of the trial, FFT will collect names, role and contact details of a key contact person and the person signing the MoU when schools are recruited. In addition, they will also collect these details for up to two teachers/TAs who will support the delivery of the Tutoring with

the Lightning Squad programme. FFT will share these data with NFER, who will then contact the key project contact person to provide further information about the trial and inform them of the trial activities. During the course of the trial, NFER and FFT will share any updates to staff data with each other to ensure that their records are up-to-date.

NFER will also collect qualitative and quantitative online survey data from school staff to gather their views on the Tutoring with the Lightning Squad programme. We will collect additional qualitative data on teachers'/TAs' attitudes towards the programme via interviews.

Pupil Data

After signing the MoU, schools approve the connection of their Management Information System (MIS) to FFT platforms. Schools will be able to indicate which pupils in Years 3 and 4 have been nominated for screening and the final 8 pupils from each year group eventually selected to participate in the trial. While FFT will receive more comprehensive data for these pupils through the connection with the school's MIS, FFT will share the following pupil data with NFER:

- Unique Pupil Number (UPN)
- First Name
- Last Name
- Date of birth
- Year group
- Scores from the Reading Assessment Programme
- Tutoring with the Lightning Squad platform usage data (only for pupils receiving the intervention)

All personal data will be shared via secure, password-protected data sharing portals.

In addition to the above, NFER will access the following data for all pupils participating in the trial from the DfE's National Pupil Database (NPD):

- FSM status
- EAL (whether the pupil speaks English as an Additional Language)
- SEND (Special Education Needs and Disability status)
- Gender
- Ethnicity
- Absence rate

To obtain the information from the NPD, NFER will provide the Data Sharing Team at the DfE with the names of the pupils, their dates of birth and UPNs, allowing a match to NPD.

NFER will collect the following pupil data directly from schools:

- Year 4 pupils' attitudes to reading using the Progress in International Reading Literacy Study (PIRLS) questionnaire at baseline and endpoint.
- Reading attainment measured using the WRMT-III at endpoint for participating pupils in Years 3 and 4.

As part of the IPE, NFER will conduct online surveys of SCs, teachers/TAs and pupils and observations of SC and tutor training sessions. A small number of schools will also be invited to participate in pupil focus groups (these will be audio recorded) and interviews with SCs and tutors. These case study visits will also include observation of a TWL session, which will be video recorded. All NFER staff visiting schools will have up-to-date DBS checks. All data gathered during interviews will be stored securely. No names of individuals will be used in any report arising from this work.

Within three months of the end of project, NFER will send school and pupil data to a separate team at FFT, which is also EEF's data archive partner. At this point, EEF's data archive partner will keep a copy of the data and EEF will become the Data Controller. NFER will retain personal data for one year after report publication in case there are any queries about the report. One year after the report publication (expected to be May 2027), all personal data will be securely deleted.

NFER and FFT are independent data controllers. NFER is the independent data controller for the evaluation, which means that NFER determines the purposes of data processing for the evaluation. FFT is the independent data controller for programme delivery and is also the data processor for the purposes of the evaluation.

Personnel

Name	Organisation	Role and Responsibilities
Programme Management		
Faizaan Sami	EEF	Evaluation Manager – key point of contact for NFER
Ben Sillitoe	EEF	Programme Manager – key point of contact for FFT
Evaluation Team		
Stephen Welbourne	NFER	Project Director – strategic leadership and responsible for overall delivery of the trial
Aarti Sahasranaman	NFER	Trial Manager – day-to-day management of the trial and delivery of the trial design, main point of contact for EEF
Lillian Flemons	NFER	IPE Lead – design and delivery of the IPE
Gemma Schwendel	NFER	Statistician
Sophie Ainsby	NFER	Research Operations Lead - overall data collection and school communications strategy
Tom Dickinson	NFER	Operations Manager – day-to-day operations including preparation of recruitment documents, coordinating data collection and point of contact for schools participating in the trial
Delivery Team		
Paul Charman	FFT	Managing Director – Project Director, provide strategic leadership
Laura James	FFT	Operations Manager – day-to-day operations including preparation of recruitment documents, coordinating recruitment processes/strategy, project management, and point of contact for EEF/NFER
Katie Kielty	FFT	Education Product Manager

Risks

S. No.	Risk	Risk Assessment		Mitigation/Counter Measures/Contingencies
		Likelihood	Impact	
1	Insufficient schools recruited by FFT (Target: 72)	Low	High	- NFER will input into recruitment material and will work closely with FFT. If required, our experienced operations team can assist with recruitment through a separate grant agreement. - Decide and monitor pre-agreed recruitment targets to identify any unfavourable trends early on to act quickly. - The design envisages that both arms will receive the intervention in different year-groups, which will make this trial more attractive to schools. - Schools with mixed year groups are eligible.
2	School attrition from trial and primary analysis	Low	Moderate	- Sign up to the trial via Memorandum of Understanding with clear identification of requirements. - Clear initial and ongoing communication with one key contact per school explaining principles, expectations, timelines and next steps. - Offer support webinar during baseline data collection to allow practitioners to ask questions. - The RAP screening test used to identify pupils suitable for TWL will also be used as the baseline measure, thus reducing the burden of baseline data collection on schools. - Schools will only get randomised if they have completed baseline testing. - Both trial arms receive the intervention reducing likelihood of post randomisation attrition. - NFER's independent Test Administrators (TAs), all ex-teachers, will serve to ensure very high follow-up rates at endline, visiting schools at a convenient time. TAs will also be present at each school for three days, which should serve to minimise pupil attrition. - Recruit 10% more schools than indicated by the power calculation. Schools receive incentive payments contingent on completing the baseline and endpoint assessments.
3	Staffing pressures mean TAs do not have time to deliver 50 – 60 programme sessions	Low	Moderate	NFER has allowed an extended period for delivery (November – May) which should ensure that all schools will be able to finish delivery before endline testing in June.

4	Contamination of the control sample through TA sharing between year groups	Moderate	Moderate	- Schools will be reminded of the importance of preventing contamination when selecting the TA to deliver the programme. - School coordinators and TAs will be asked about any possible contamination as part of the IPE data collection, and this will be taken into account when interpreting the impact findings.
5	Spillover risks due to schools delivering the intervention to the wrong year-group	Low	High	- FFT will create user accounts only for TAs and pupils in intervention year groups as per the randomisation allocation provided by NFER minimising the likelihood of spillover risks. - FFT will have sight of usage of the TWL platform and will be in constant touch with SCs throughout the duration of intervention delivery.
6	School does not have the necessary IT capacity for intervention delivery and outcome measurement	Low	Moderate	- Clear initial communication outlining IT requirements at EoI and MoU stage both in terms of WiFi connectivity and number and type of devices. - FFT will be able to supply schools with devices if required.
7	Changes to the project team due to sickness, absence or staff turnover	Moderate	Moderate	- NFER has a large research department with numerous researchers and research associates experienced in evaluation who could be redeployed. - Clear and accurate project documentation would support continuity in the event of any team changes.
8	NFER is unable to obtain the necessary analytic data from the TWL platform	Low	Moderate	- NFER will liaise closely with FFT to agree a data spec in advance and ensure regular monitoring of the analytic data available. - NFER will ask FFT to share the data in batches covering shorter periods rather than extracting it all together at the end.
9	Difficulty in securing target engagement with IPE activities	Moderate	Moderate	- Communication with schools explaining research benefits. - Ongoing reminders. - Flexibility in timings of school visits and interviews. - Close liaison with the delivery team to support IPE engagement. - Online data collection (including remote interviews) where possible to minimise burden. 'Thank you' payments of £150 for all case study schools.
10	Parents are unwilling to consent to their child's participation in data collection activities	Low	Moderate	- Parents will be provided with a clear Privacy Notice outlining the value of their child participating in the research, as well as any associated risks and how these will be mitigated. - Schools will be asked to encourage parents to allow their children to participate.

Timeline

Table 5: Timeline

Dates	Activity	Staff responsible/leading
May – October 2024	Project set-up Complete set-up meetings and IDEA workshop, finalise grant agreement	Stephen Welbourne, Aarti Sahasranaman
November 2025 – January 2025	Prepare for recruitment Finalise recruitment documents including privacy notices	Sophie Ainsby, Tom Dickinson
January – May 2025	Recruitment and pupil data collection School recruitment and MIS pupil data sharing set-up with FFT	FFT
	Trial Protocol Protocol draft submitted in March 2025	Aarti Sahasranaman
	Data sharing process Finalise data sharing agreement and data flow diagram	Sophie Ainsby, Tom Dickinson, Aarti Sahasranaman (DSA)
May 2025	School data sharing All recruited schools shared by FFT with NFER by 23 rd May 2025	FFT
June – July 2025	Pupil nomination by schools Schools select around 30 pupils each from Year 3 and Year 4 for screening	Schools
	NFER contact schools with activities for autumn term	Sophie Ainsby, Tom Dickinson
September 2025	Baseline Testing and Pupil Selection SCs attend online workshop about RAP Nominated pupils in years 3 and 4 complete baseline test/screener (RAP) Schools select 8 pupils from each year and put them forward for the trial Final 8 selected pupils from Year 4 complete baseline pupil survey SCs complete online survey	Schools, FFT

October 2025	Randomisation NFER informs FFT and schools of random allocation	Gemma Schwendel, Tom Dickinson (notify schools and FFT)
	Intervention pupils complete initial assessment on TWL platform; teachers pair pupils of similar abilities	Schools
End October/Early November 2025	Training SC and two teachers/TAs complete day-long online training	FFT, Schools
November 2025 – May 2026	TWL intervention delivery	FFT
	IPE case studies	
January - April 2026	Statistical Analysis Plan (SAP) first draft submitted in January '26; final SAP published in April '26 NPD application submitted (only if background pupil data from FFT does not meet our quality threshold)	Gemma Schwendel
June - July 2026	Endpoint Testing WRMT-III assessment (Years 3 and 4) and endline pupil survey (only Year 4) completed in schools SCs and Teaching Assistants complete online survey	
August – November 2026	IPE data analysis (starts August) Impact data analysis (starts September)	
Jan 2027	First draft of evaluation report	
May 2027	Evaluation report published	
July 2027	Data archived	

References

- Ashraf, B., Singh, A., Uwimpuhwe, G., Higgins, S. and Kasim, A. (2021) 'Individual participant data meta-analysis of the impact of educational interventions on pupils eligible for Free School Meals', *British Educational Research Journal*, 47(6), pp. 1675–1699. Available at: <https://doi.org/10.1002/berj.3749>.
- Bibby, A., Charman, P., Fischer, M. and Hilton, E. (2023) *FFT Tutoring Data Impact Report: 2021-2022*. FFT. Available at: <https://www2.fft.org.uk/tutoring-data-impact-report> (Accessed: 29 April 2025).
- Bulus, M., Dong, N., Kelcey, B. and Spybrook, J. (2021) 'PowerUpR: power analysis tools for multilevel randomized experiments'. Available at: <https://cran.r-project.org/web/packages/PowerUpR/index.html> (Accessed: 6 March 2025).
- Castles, A., Rastle, K. and Nation, K. (2018) 'Ending the reading wars: reading acquisition from novice to expert', *Psychological Science in the Public Interest*, 19(1), pp. 5–51. Available at: <https://doi.org/10.1177/1529100618772271>.
- Chambers, B., Slavin, R.E., Madden, N.A., Abrami, P., Logan, M.K. and Gifford, R. (2011) 'Small-group, computer-assisted tutoring to improve reading outcomes for struggling first and second graders', *The Elementary School Journal*, 111(4), pp. 625–640. Available at: <https://doi.org/10.1086/659035>.
- Chen, J., Wang, M., Kirschner, P.A. and Tsai, C.-C. (2018) 'The role of collaboration, computer use, learning environments, and supporting strategies in cscl: a meta-analysis', *Review of Educational Research*, 88(6), pp. 799–843. Available at: <https://doi.org/10.3102/0034654318791584>.
- Cheung, A.C.K. and Slavin, R.E. (2013) 'Effects of educational technology applications on reading outcomes for struggling readers: a best-evidence synthesis', *Reading Research Quarterly*, 48(3), pp. 277–299. Available at: <https://doi.org/10.1002/rrq.50>.
- Decker, S.L., Strait, J.E., Roberts, A.M. and Wright, E.K. (2018) 'Cognitive mediators of reading comprehension in early development', *Contemporary School Psychology*, 22(3), pp. 249–257. Available at: <https://doi.org/10.1007/s40688-017-0127-0>.
- Department for Education (2023) *The reading framework*. Available at: https://assets.publishing.service.gov.uk/media/664f600c05e5fe28788fc437/The_reading_framework_.pdf (Accessed: 29 April 2025).
- DfE (2024) *Schools, pupils and their characteristics, academic year 2023/24*, GOV.UK. Available at: <https://explore-education-statistics.service.gov.uk/find-statistics/school-pupils-and-their-characteristics> (Accessed: 6 March 2025).
- DfE and STA (2018) *National curriculum assessments at Key Stage 2 in England, 2018 (provisional)*, GOV.UK. Available at: <https://www.gov.uk/government/statistics/national-curriculum-assessments-key-stage-2-2018-provisional/national-curriculum-assessments-at-key-stage-2-in-england-2018-provisional--2> (Accessed: 4 March 2025).

Education Standards Research Team (2012) *Research evidence on reading for pleasure*. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/284286/reading_for_pleasure.pdf#:~:text=Evidence%20suggests%20that%20there%20is%20a%20positive%20relationship,success%20than%20their%20family%E2%80%99s%20socio-economic%20status%20%28OECD%2C%202002%29. (Accessed: 11 February 2025).

EEF (2021a) *Collaborative learning approaches*. Available at: <https://educationendowmentfoundation.org.uk/education-evidence/teaching-learning-toolkit/collaborative-learning-approaches> (Accessed: 4 March 2025).

EEF (2021b) *Small group tuition*. Available at: <https://educationendowmentfoundation.org.uk/education-evidence/teaching-learning-toolkit/small-group-tuition> (Accessed: 4 March 2025).

EEF (2022a) *Implementation and process evaluation guidance for EEF evaluations*. Available at: <https://d2tic4wvo1iusb.cloudfront.net/production/documents/evaluation/evaluation-design/EEF-IPE-Guidance-August-2022.pdf?v=1739885156> (Accessed: 16 April 2024).

EEF (2022b) *Statistical analysis guidance for EEF evaluations*, p. 6/3/25. Available at: <https://d2tic4wvo1iusb.cloudfront.net/production/documents/evaluation/evaluation-design/EEF-Analysis-Guidance-Website-Version-2022.14.11.pdf?v=1709155078>.

EEF (2023) *Cost evaluation guidance for EEF evaluations*. Available at: https://d2tic4wvo1iusb.cloudfront.net/production/documents/evaluation/evaluation-design/Cost-Evaluation-Guidance-Feb_2023.pdf?v=1739793629 (Accessed: 26 February 2025).

Gibbons, S., Scrutiny, V. and Telhaj, S. (2018) *Teacher turnover: does it matter for pupil achievement? CEP discussion paper no. 1530*. CEP discussion paper no. 1530. Centre for Economic Performance. Available at: <https://eric.ed.gov/?id=ED583870> (Accessed: 6 February 2025).

GL Assessment (2020) *Read all about it: why reading is key to GCSE success*. London: GL Assessment. Available at: <https://reports.gl-assessment.co.uk/whyreading/images/gl-assessment-report-gcse-success.pdf?v=3.5> (Accessed: 6 February 2025).

GOV.UK (2025) *Key Stage 2 attainment, academic year 2023/24*. Available at: <https://explore-education-statistics.service.gov.uk/find-statistics/key-stage-2-attainment> (Accessed: 4 March 2025).

Harrison, W. and Higgins, S. (2023) *A global evidence review for policymakers: findings from meta-analyses of tutoring*. Available at: https://edtlive.b-cdn.net/live/media/25macv0c/edt_global-evidence-review_working-paper-v3_0723.pdf (Accessed: 4 March 2025).

Hernandez, D.J. (2012) *Double jeopardy: how third-grade reading skills and poverty influence high school graduation*. Annie E. Casey foundation. Available at: <https://assets.aecf.org/m/resourcedoc/AECF-DoubleJeopardy-2012-Full.pdf> (Accessed: 4 March 2025).

HM Government (2022) *Levelling up the United Kingdom: missions and metrics - technical annex*. CP 604. Available at: https://assets.publishing.service.gov.uk/media/620b772be90e0710a7b3ffca/Technical_annex_-_missions_and_metrics.pdf (Accessed: 4 March 2025).

IEA (2020) 'Progress in International Reading Literacy Study: Student Questionnaire - Grade 4'. Available at: https://pirls2021.org/wp-content/uploads/2023/05/P21_StudentQuestionnaire.pdf (Accessed: 13 February 2025).

Joffe, H. (2012) 'Thematic Analysis', in D. Harper and A.R. Thompson (eds) *Qualitative research methods in mental health and psychotherapy: a guide for students and practitioners*. Chichester: John Wiley & Sons, Ltd., pp. 209–223.

Johnson, D.W. and Johnson, R.T. (1975) *Learning together and alone: cooperation, competition and individualization*. Englewood Cliffs, New Jersey: Prentice-Hall, Inc. Available at: <https://www.jstor.org/stable/43763555> (Accessed: 13 February 2025).

Lewin, C., Morris, S., Ainsworth, S., Gellen, S. and Wicker, K. (2024) *Peer assisted learning strategies – UK (PALS-UK): a whole-class reading approach*. Available at: https://d2tic4wvo1iusb.cloudfront.net/production/documents/projects/PALS_report_final.pdf?v=1740998304 (Accessed: 4 March 2025).

Lewin, C., Smith, A., Morris, S. and Craig, E. (2019) *Using digital technology to improve learning: evidence review*. Available at: https://d2tic4wvo1iusb.cloudfront.net/production/documents/guidance/Using_Digital_Technology_to_Improve_learning_Evidence_Review.pdf?v=1708943413 (Accessed: 25 February 2025).

Madden, N.A. and Slavin, R.E. (2017) 'Evaluations of technology-assisted small-group tutoring for struggling readers', *Reading & Writing Quarterly*, 33(4), pp. 327–334. Available at: <https://doi.org/10.1080/10573569.2016.1255577>.

Marguerita, L. and Conlon, G. (2016) *The impact of literacy, numeracy and computer skills on earnings and employment outcomes*. OECD Education Working Papers, No. 129. Available at: https://www.oecd.org/content/dam/oecd/en/publications/reports/2016/03/the-impact-of-literacy-numeracy-and-computer-skills-on-earnings-and-employment-outcomes_g17a278d/5jm2cv4t4gzs-en.pdf (Accessed: 29 April 2025).

MAXQDA (2022) *Document and code variables in MAXQDA*. Available at: <https://www.maxqda.com/help-mx22/variables/document-code-variables-maxqda> (Accessed: 17 February 2025).

MAXQDA (2025) *Intercoder agreement*. Available at: <https://www.maxqda.com/help-mx22/teamwork/problem-intercoder-agreement-qualitative-research> (Accessed: 17 February 2025).

Merewood, J., Senator, B., Hutton, E., Reed, M., Subosa, M. and Brown, E.R. (2022) *Independent evaluation of the Lexia® Core5® Reading Programme: A cluster randomised controlled trial Evaluation Protocol*. Available at: <https://d2tic4wvo1iusb.cloudfront.net/production/documents/projects/Evaluation-protocol-Lexia.pdf?v=1725959609> (Accessed: 6 February 2025).

Mulcahy, E., Bernardes, E. and Baars, D.S. (2019) *The relationship between reading age, education and life outcomes*. Available at: <https://cfey.org/wp-content/uploads/2019/03/The-relationship-between-reading-age-education-and-life-outcomes.pdf> (Accessed: 29 April 2023).

Mullis, I.V.S. and Martin, M.O. (2019) *PIRLS 2021 assessment frameworks*. Available at: https://pirls2021.org/wp-content/uploads/sites/2/2019/04/P21_Frameworks.pdf (Accessed: 16 April 2024).

Ross, S.M., Laurenzano, M. and Madden, N.A. (2017) *An evaluation of the lightning squad computer-assisted small group tutoring program on the reading achievement of disadvantaged students in grades 1-3*. Available at: <https://www.successforall.org/wp-content/uploads/2018/05/Ross-Laurenzano-Madden-2017.pdf> (Accessed: 4 March 2025).

Savolainen, H., Ahonen, T., Aro, M., Tolvanen, A. and Holopainen, L. (2008) 'Reading comprehension, word reading and spelling as predictors of school achievement and choice of secondary education', *Learning and Instruction*, 18(2), pp. 201–210. Available at: <https://doi.org/10.1016/j.learninstruc.2007.09.017>.

Sharples, J., Webster, R. and Blatchford, P. (2021) *Making best use of teaching assistants*. Guidance Report. London: EEF. Available at: https://d2tic4wvo1iusb.cloudfront.net/production/eef-guidance-reports/teaching-assistants/TA_Guidance_Report_MakingBestUseOfTeachingAssistants-Printable_2021-11-02-162019_wsqd.pdf?v=1738171905 (Accessed: 6 February 2025).

Singh, A., Uwimpuhwe, G., Vallis, D., Akhter, N., Coolen-Maturi, T., Higgins, S., Einbeck, J., Culliney, M. and Demack, S. (2023) *Improving power calculations in educational trials*. Available at: https://d2tic4wvo1iusb.cloudfront.net/production/documents/evaluation/methodological-research-and-innovations/Work_Package_2023-WP6_18_09_2023_FINAL.pdf?v=1741258918 (Accessed: 6 March 2025).

Slavin, R.E., Lake, C., Davis, S. and Madden, N.A. (2011) 'Effective programs for struggling readers: a best-evidence synthesis', *Educational Research Review*, 6(1), pp. 1–26. Available at: <https://doi.org/10.1016/j.edurev.2010.07.002>.

VERBI Software (2021) 'MAXQDA 2022'. Berlin: VERBI Software.

Woodcock, R. (2011) *WRMT-III - woodcock reading mastery tests, third edition*. Available at: https://www.pearsonclinical.co.uk/store/ukassessments/en/Store/Professional-Assessments/Academic-Learning/Woodcock-Reading-Mastery-Tests%2C-Third-Edition/p/P100009264.html?srsId=AfmBOorLJ3OK_MmceAfzG3_LycJU9zUsSn3RGuEP-6Wr06mtgZh4h7BV (Accessed: 10 September 2024).

Appendix A: questions from section R7 of PIRLS

- a. I like talking about what I read with other people
- b. I would be happy if someone gave me a book as a present
- c. I think reading is boring
- d. I would like to have more time for reading
- e. I enjoy reading
- f. I learn a lot from reading
- g. I like to read things that make me think
- h. I like it when a book helps me imagine other worlds

Appendix B – Sample Size calculations syntax

```
#
=====
=====

# Script Name: EELSE power calculations protocol.R =

# Script Purpose: Perform power calculations for EELS for protocol drafting =

# =

# Author: Gemma Schwendel =

# Date Created: 11/04/2023 =

# Notes: =

#
=====
=====

# =

# Revision Log =

#
=====
=====

# Date | Reason for revision

# 18/07/2024 | Stephen Welbourne Revised power calcs for 2nd setup based on 2
yeargroups and 8 pupils per school 2 form entry schools

# https://d2tic4wvo1iusb.cloudfront.net/production/documents/evaluation/methodological-
research-and-innovations/Work\_Package\_2023-WP6\_18\_09\_2023\_FINAL.pdf?v=1696233531 =

# 21/02/2025 | Gemma Schwendel Updated to reflect feedback whilst writing protocol:

# assume no school level attrition between randomisation and endline =

# and calculate with and without pupil-level attrition of 15% =

#
=====
=====
```

```
library(PowerUpR)
```

```
library(ggplot2)
```

```
rm(list=ls())
```

```
##### Assumptions
```

```
#####
```

```
### The model used is a 3-level block randomisation with random effects across level 3 blocks
```

```
## Level One: Pupil
```

```
## Level Two: Year Group
```

```
## Level Three: School
```

```
### Global Assumptions
```

```
power = 0.8
```

```
alpha = 0.05
```

```
ICC = 0.115 ## ICC KS2 English for NPD whole
```

```
ICC2 = ICC*(2/3) # Assumed Overall ICC 0.11 with 2/3 accounting for variation at class level
```

```
ICC3 = ICC*(1/3) # Assumed Overall ICC 0.11 with 1/3 accounting for variation at school level
```

```
om = 0.5 # Assumed treatment effect heterogeneity = 0.5
```

```
ppn = 0.5 # Assume half go into each treatment arm
```

```
g3 = 1 # number of covariates at level 3. Assumed one. Varying this has little difference
```

```
r21 = 0.662^2 # proportion of level 1 variance in the outcome explained by level 1 covariates
```

```
r22 = 0 # proportion of level 2 variance in the outcome explained by level 2 covariates: Assumed  
no level 2 covariates
```

```
r2t3 = 0.6^2 # proportion of treatment effect variance among level 3 units explained by level 3  
covariates
```

```
n = 8 # Eight pupils per class
```

```
cls = 2 # Two classes per year group
```

J = 2 # Year-groups per school

K = 72 # Number of schools

Without attrition

pupil_ret = 1.0 # As per Faizaan's request, calculate with no attrition

school_ret = 1.0 # Assumed all schools are retained

ppn_fsm = 0.39 # 39% of those not meeting expected standard are FSM eligible

ppn_eal = 0.25 # 25% of those not meeting expected standard have EAL

ppn_sen = 0.4 # 40% of those not meeting expected standard have SEN

All targeted pupils

mdes.bcra3r2(power=power, alpha=alpha, two.tailed=TRUE, rho2=ICC2, rho3=ICC3,
omega3=om,

p=ppn, g3=g3, r21=r21, r22=r22, r2t3=r2t3, n=n*pupil_ret, J=J,
K=K*school_ret)\$mdes[1,1]

FSM pupils within targeted pupils

mdes.bcra3r2(power=power, alpha=alpha, two.tailed=TRUE, rho2=ICC2, rho3=ICC3,
omega3=om,

p=ppn, g3=g3, r21=r21, r22=r22, r2t3=r2t3, n=n*ppn_fsm*pupil_ret, J=J,
K=K*school_ret)\$mdes[1,1]

And with 15% pupil attrition, no school attrition

pupil_ret = 0.85

school_ret = 1.0 # Assumed all schools are retained

All targeted pupils

mdes.bcra3r2(power=power, alpha=alpha, two.tailed=TRUE, rho2=ICC2, rho3=ICC3,
omega3=om,

```
p=ppn, g3=g3, r21=r21, r22=r22, r2t3=r2t3, n=n*pupil_ret, J=J,  
K=K*school_ret)$mdes[1,1]
```

```
# FSM pupils within targeted pupils
```

```
mdes.bcra3r2(power=power, alpha=alpha, two.tailed=TRUE, rho2=ICC2, rho3=ICC3,  
omega3=om,
```

```
p=ppn, g3=g3, r21=r21, r22=r22, r2t3=r2t3, n=n*ppn_fsm*pupil_ret, J=J,  
K=K*school_ret)$mdes[1,1]
```

Appendix C – sensitivity analysis of ICC split

Proportion of ICC at year group level	Proportion of ICC at school level	MDES for all targeted children
10%	90%	0.142 (0.151)
20%	80%	0.150 (0.158)
30%	70%	0.157 (0.164)
33%	67%	0.159 (0.167)
40%	60%	0.163 (0.171)
50%	50%	0.170 (0.177)
60%	40%	0.176 (0.183)
67%**	33%**	0.180 (0.187)
70%	30%	0.182 (0.189)
80%	20%	0.188 (0.195)
90%	10%	0.194 (0.200)

* Note that the MDES values here assume no pupil-level attrition and (15%) pupil-level attrition, in line with the values presented in Table 2

** This is the ICC split assumed in Table 2